

Landslide Susceptibility Mapping Using Logistic Regression Analysis and GIS Tools

**Abolghasem Akbari, Fadzil Bin Mat Yahaya,
Mahmoud Azamirad, Mohsen Fanodi**

Corresponding author: Abolghasem Akbari

e-mail: akbariinbox@yahoo.com, Tel: +6095492983 Fax: +6095492998

*Address: Faculty of Civil Engineering & Earth Resources, University Malaysia
Pahang Lebuhraya Tun Razak 26300 Gambang, Kuantan, Pahang, Malaysia*

ABSTRACT

Landslide is a major threat in many regions with humid climate condition. In recent years, this phenomenon has been accelerated by human activities mainly by rural and urban development projects. This research integrates the GIS tools and multivariate regression analysis for landslide susceptibility modeling (LSM) in north of Iran. To map the landslide susceptibility, ten potential independent variables were selected as effectual factors, including geological formation, terrain elevation, terrain slope and aspect, proximity to roads, proximity to faults and proximity to main rivers, soil unit, land use and annual rainfall. A GIS-database was developed containing all variables for the study area. Previous records of landslides in the study area were mapped based on inventory reports, satellite image processing and field survey using handheld GPS. The slope, proximity to roads, elevation, aspect and soil units was found to be the most effective factors in landslides respectively. Five other factors had no significant effect on landslides in this region. Landslide susceptibility map was then generated based on multivariate regression equation in a raster GIS environment and classified in five susceptibility classes. About 11.16% of the study area has very low susceptibility, 40.36% has low susceptibility, 32.37% has moderate susceptibility, 12.90 % has high susceptibility and 3.23 % has very high susceptibility.

KEYWORDS: Landslide, Multivariate regression, GIS, SPSS

BACKGROUND

Preventing natural disaster such as landslide is one of the best practices in watershed management activities. Susceptibility map provides a document that describes the likelihood or possibility of new landslides occurring in an area, and therefore helping to reduce future potential damages in future. Depending on the landform, several factors can cause or accelerate the Landslide. According to the previous researches, Human activities (Ayalew and Yamagishi, 2005), land morphology (Gorsevski et al., 2006), soil characteristics Regmi et al., 2010, slope (Lee, 2005; Yalcin, 2005), aspect (Lee et al., 2004; Yalcin, 2008) and climate conditions (Suzan and Doyuran, 2004; Komac, 2006), proximity to some watershed features such as rivers and fault (Ayalew and Yamagishi, 2005; Yalcin, 2005) are the most important parameters. Landslide susceptibility modeling (LSM) and analysis is done through varieties of methods and techniques. A detailed outline of the various methods and their advantages and disadvantages are systematically compared in literature (van Westen et al. 2006; Keefer and Larsen 2007). GIS is an effective tool for managing

and manipulating the spatial data with an appropriate model for mapping landslide susceptibility. Statistical model like logistic regression (Ayalew and Yamagishi 2005; Duman et al. 2006; Nefeslioglu et al. 2008; Pradhan 2010; Ercanoglu and Temiz 2011), frequency ratio (Lee and Sambath 2006; Lee and Pradhan 2007; Vijith and Madhu 2008; Yilmaz 2010; Constantin et al. 2011), and certainty factor (Lan et al. 2004; Fenghau et al. 2010; Kanungo et al. 2011) are successfully used to map landslide susceptibility. Due to The natural variability of the geotechnical parameters and the uncertainties concerning the boundary conditions, statistical and probabilistic approaches are more in favor. On the other hand the principal parameters are distributed statistically to account for their spatial variability. However, sufficient and accurate information about the landslide and contributing parameters are needed to construct lanslid prediction model (Zhu and Huang 2006).

The multivariate approach: Logistic Regression (LR)

It is believed that among the wide range of statistical methods proposed in landslide susceptibility mapping , LR analysis has proven to be one of the most reliable approaches (Ayalew and Yamagishi, 2005; Chau and Chan, 2005; Lee and Sambath, 2006; Lee and Pradhan, 2007; Rickli and Graf 2009; Sujatha et al. 2012 and Chen and Wang, 2007). Basically, LR analysis relates the probability of landslide occurrence (having values from 0 to 1) to the “logit” Z (where $-1 < Z < 0$ for higher odds of non-occurrence and $0 < Z < 1$ for higher odds of occurrence). In the LR formulae, the probability of landslide occurrence is expressed by

$$P_r = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}} \quad \text{Eq.1}$$

The logit Z is assumed to contain the independent variables on which landslide occurrence may depend. The LR analysis assumes the term Z to be a combination of the independent set of geographical variables X_i ($i=1,2,\dots,n$) acting as potential causal factors of landslide phenomena. The term Z is expressed by the linear form

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad \text{Eq.2}$$

where coefficients i ($i=1,2,\dots,n$) are representative of the contribution of single independent variables X_i to the logit Z and 0 is the intercept of the regression function. It must be noted that the LR approach does not require, or assume, linear dependencies between dependent term of P_r and the independent set of variables representing causal factors. An exponential equation and coefficients are estimated using maximum likelihood criterion and correspond to the estimation of the more likely unknown factors. Multivariate regression analysis plays a central role in statistics that cause one of the most powerful and commonly used techniques (McCullagh and Nelder, 1989). However, this approach inherently has some limitations such as generalizing and simplifying of causal factors. In addition, this method does not take into account the temporal aspects of landslides and is not able to predict the impact of changes in the controlling conditions.

METHODOLOGY

Study area

The study area is located on north of Iran and close to Caspian Sea (see Fig. 1). The shape of the study area is extended North-East Ward and covers about 62.07 km². Geographically this area expand in Latitude 53° 00' 10"N to 53° 06' 35"N and longitude 36° 20' 50"E to 36° 30' 50"E. According to Khazar.C.C, (1985) the annual average rainfall is about 650 mm.

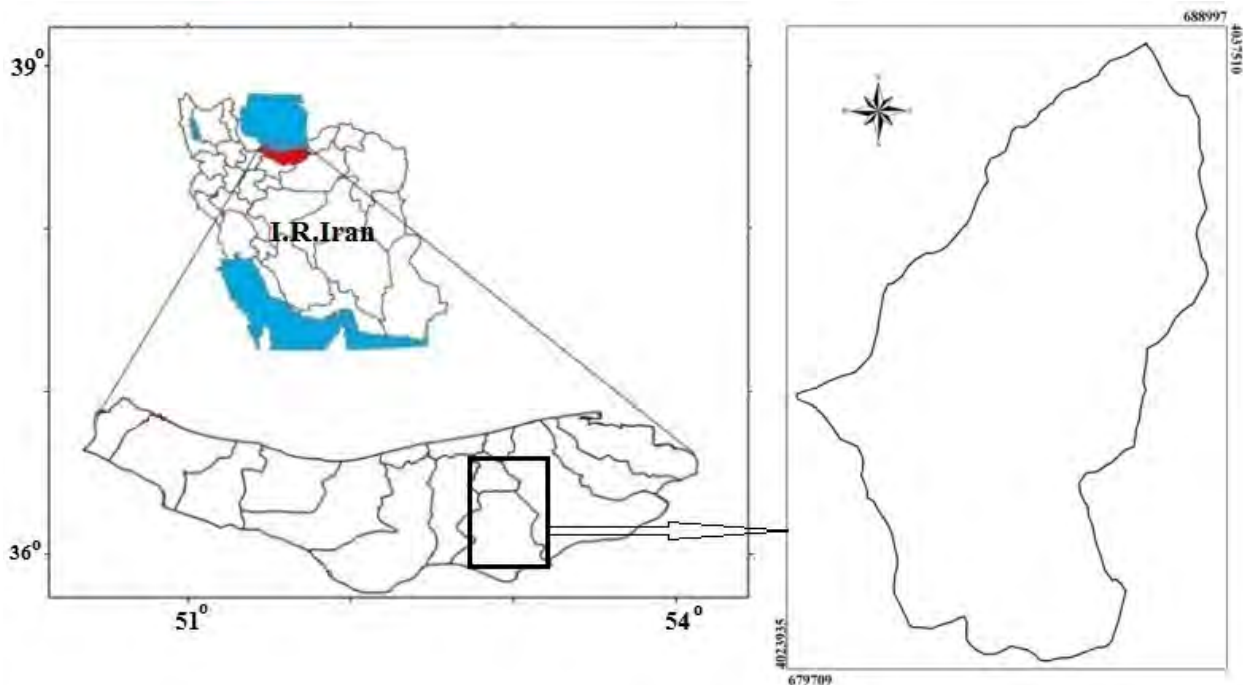


Figure 1: Study area

Materials

This research employs GIS tools and multivariate regression analysis for LSM. To achieve the objective of the study, a geodatabase including all investigated factors was generated. The affected area by landslides was mapped by preliminary investigation on Landsat ETM⁺ and IRS images, filed survey and piking the positions using handheld GPS Map 76 scx . The Soil unit map of the study area was obtained from the local organization at scale of 1:25000, which contains four different units which has been calssified in four class including 1.5.2, 1.5.3, 2.5.2 and 2.5.4. were then weighted based on depth and drainage conditions (see Figure 2.1). Digital Elevation Model (DEM) was generated with 10 meters cell size from digital topo map of the study area at scale 1:25000. DEM was classified into four classes with domains of 0-291, 291-434, 434-592, 592-890 meters (see Figure 2.2). Terrain slope map was derived from DEM in four classes. Slope map was derived from DEM and classified in four classes based on practical applications and land suitability (see Figure 2.3). Terrain Aspect map was also derived from DEM and classified in five classes including north, south, east, west and Flat (see Figure 2.4). To investigate the impact of proximity to roads three type of buffer zone range from 100 to 300 meter were created for existing road map extracted from topographic map at scale 1:25000 (see Figure 2.5). The same way six buffer zones with 100 m intervals were created to account the influence of proximity to faults (see Figure 2.6). Domain classes include 0-100, 100-200, 200-300, 300-400, 400-500 and greater than 500 meters. Similarly, six buffer zones with 100 m interval were generated to investigate the influence of proximity to main rivers with domain class of 0-100 m, 100-200 m, 200-300 m, 300-400 m, 400-500 and greater than 500 meters (see Figure 2.7). Spatial distribution of Rainfall considerably varies from lower lands to upper mountains of the study area. Based on previous studies (Khazar.C.C, 1985) mean annual precipitation ranges from 417 mm in lowlands to 872 mm in highlands. Spatial distribution of rainfall was determined from existing empirical equation (see Eq. 2) that relates elevation to the mean annual rainfall over the study area. For this purpose DEM was employed to generate areal distribution of annual rainfall in the study area (see Figure 2.8). Land use of the study area is almost

homogenous and covered with forest. Land use of the study area was taken from natural resource organization of Mazandaran's province (See figure 2.9). Geological formation of the study area was determined from geology maps provided by National Geological Organization at scale 1:100000. The permanent lithology of the study area includes M2.3m.s.l, Plqc.s and Q2 that are very susceptible to landslide (see Figure 2.10). Based on the nature of the study area of geological formation is almost homogenous and covered with M2.3m.s.l.

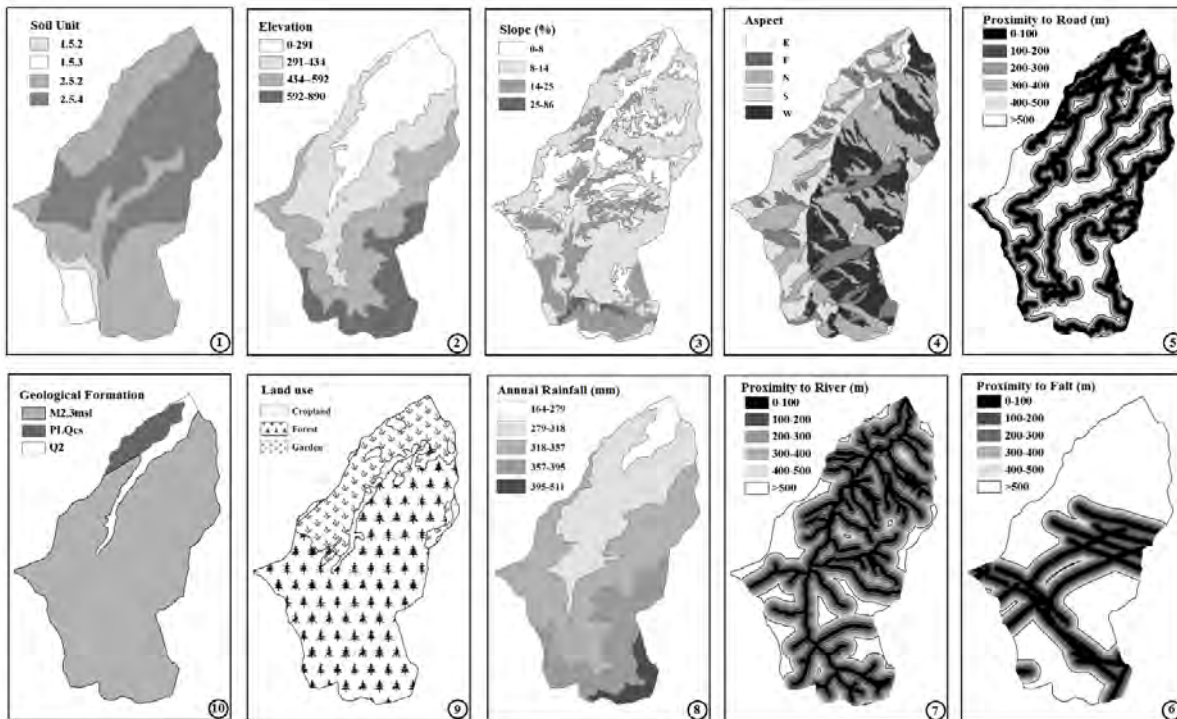


Figure 2: GIS-database investigated for mapping the landslide susceptibility including: 2.1: Soil Units, 2.2: Elevation, 2.3: Slope, 2.4: Aspect, 2.5: Proximity to road, 2.6 Proximity to fault, 2.7 Proximity to river, 2.8: Annual Rainfall, 2.9: Landuse, 2.10 Geological Formation

$$\text{Rainfall}_{\text{annual}} = 323.2 \log \times \text{DEM} - 53.2$$

Eq.3

where $\text{Rainfall}_{\text{annual}}$: is annual rainfall (mm) and DEM is Elevation (m)

Homogenous units

Statistical analysis on observed landslides (OLs) indicated that three factors including proximity to rivers and faults and annual rainfall have no significant impact on this area. Therefore they were eliminated from further analysis. Homogeneous units were identified by overlaying the remaining independent variables in the form of vector layers. Then, the weight factors for each layer were calculated by dividing area of OLs to the area of homogeneous units. Weight factors were transferred to the quantitative values from zero to 10. As shown in table 1, factors were coded from 1 to 5. By overlaying the seven investigated factors, 158 homogeneous units were identified.

Table 1: Major factors and coding approach for homogeneous units

No	variables	Code				
		1	2	3	4	5
1	Soil Unit	1.5.2	1.5.3	2.5.2	2.5.4	
2	Terrain Elevation	0-291	291-434	434-592	592-890	
3	Terrain Slope	0-8	8-14	14-25	>25	
4	Terrain Aspect	<i>N</i>	<i>E</i>	<i>S</i>	<i>W</i>	<i>F</i>
5	Geological Formation	Q ₂	PLQ _{CS}	M _{2,3} m.s.l		
6	Proximity to Road	0-100	100-200	200-300	>300	
7	Land Use	Forest	Cropland	Garden		

Prediction of landslide susceptibility

Multivariate regressions method in SPSS software was used with seven factors including landuse, formation, Soil unit, elevation, slope, aspect and proximity to road against 158 homogeneous units. In this method, independent variables were entered based on its correlation with depend variable which is landslide susceptibility. Variables with greater correlation are entered earlier and variables with lower correlation were eliminated. The landslide prediction model which is a numerical presentation of LSM was formed as shown in Eq.3 and Eq.4. In final step, Raster-base LSM was classified into five quantitative classes including Very low Susceptibility (VLS); low Susceptibility (LoS), moderate Susceptibility (MS), high Susceptibility (HS) and very high Susceptibility (VHS) (see Figure 3).

$$Z = -0.093 + 0.005TS + 0.003PR + 0.004TE + 0.004TA + 0.002SU \quad \text{Eq.3}$$

$$LS_{Pre} = \frac{1}{1 + e^{(0.093 - 0.005TS - 0.003PR - 0.004TE - 0.004TA - 0.002SU)}} \quad \text{Eq.4}$$

where *e*: is Napierian logarithm, *LS_{Pre}*: is predicted landslide susceptibility, *TS*: is Terrain Slope, *PR*: is Proximity to Road, *TE*: Terrain Elevation, *TA*: is Terrain Aspect and *SU*: is Soil Unit.

All GIS analysis was performed with ILWIS which is public domain raster-based GIS software. however the SPSS statistical package was employed for the LR analysis.

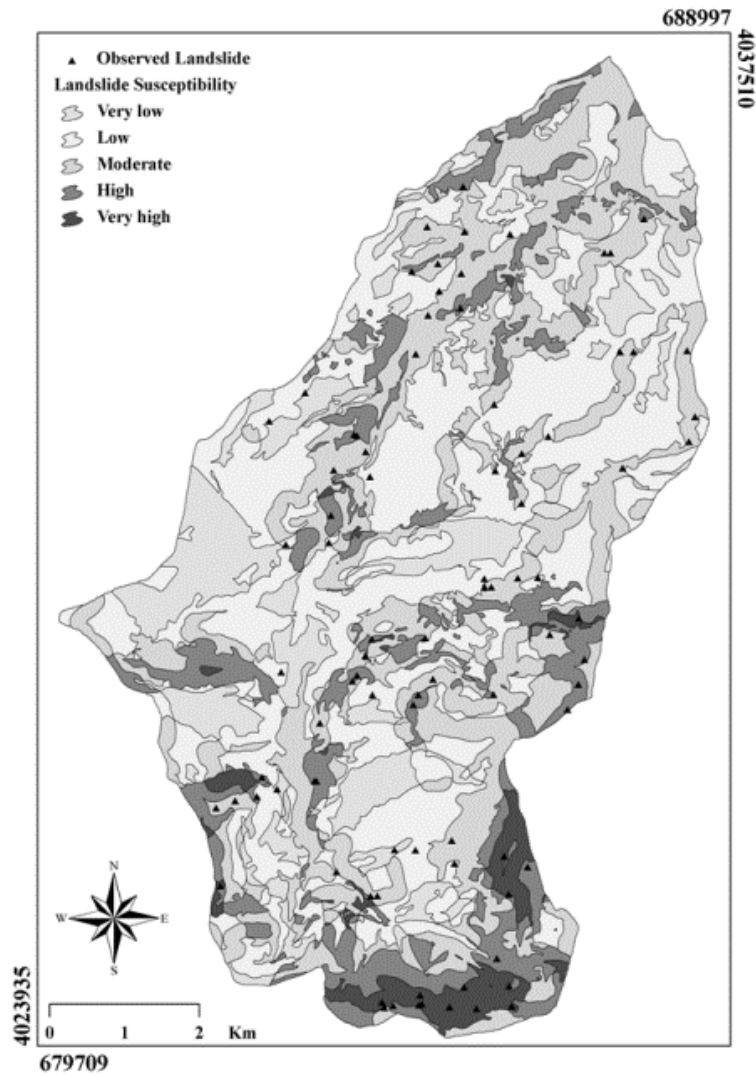


Figure 3: Landslide susceptibility map of the study area

Landslide susceptibility map was classified into five classes with equal interval as shown in Table 2.

Table 2: susceptibility classes based on Y, value.

No	Susceptibility classes	Susceptibility index
1	VLS	$LS < -0.0015$
2	LoS	$-0.0015 < LS < 0.01$
3	MS	$0.01 < LS < 0.036$
4	HS	$0.036 < LS < 0.061$
5	VHS	$0.061 < LS < 0.087$

Validation of landslide susceptibility map

To validate the accuracy of the landslide susceptibility map, observed landslides and susceptibility map was compared. As it evident in Table 3, high percentage of the landslides has been occurred in very high susceptibility area that covers the lower percentage of study zone.

Table 3: comparison of the landslide suitability classes based on observed and predicted percentage of landslide

Susceptibility classes	$Area_{pre}$ (%)	$Area_{obs}$ (%)
VLS	11.16	6.04
LoS	40.36	15.53
MS	32.37	22.87
HS	12.90	26.40
VHS	3.23	29.16

DISCUSSION

In this study area, most of observed landslides have been occurred on slopes between 14 % and 25 %. There are no any observed landslide in slopes greater than 25 percent. It is because of the fact that in very steep slopes the soil depth and its load are decreased (Knopen et al, 2006). Most of landslides have occurred in north and west aspects, which is due to angle of sun radiation and soil moisture condition . Usually, in Iran the run way of wet air masses is in north-west direction. In addition, North side slopes is not exposed to radiation, therefore it is wetter than other aspects. The rate of landslide is increased with in higher elevations. However, in elevation range from zero to 291 meters (fist cluster of elevation) the rate of landslide is more than clusters 2 and 3. It may relate to landuse change which is crop in lower lands to forest in upper lands. According to Yalcin (2008), geology formation may have significant influences on landslide; however, in this area geology formation was not significant, because there is no much diversity in formation. Analysis of observed landslides indicated that most of events have occurred near or close to the roads. This may due to development activities and constructing roads with deep trenches. It is also observed that some notable landslides have occurred in soil unit 2.5.2; it may be due to low infiltration capacity of this zone, which is a barrier for drainage system. Whereas, any landslide have been recorded in soil unit of 1.5.3. It is may be due to gentler slope and higher infiltration capacity of this unit.

CONCLUSION

In this research ten factor were investigated to map the landslide susceptibility. It was found out that five factors including slope, aspect, and proximity to road, elevation and land type are main casual factors of landslides in the study area. It was also demonstrate the impact of human activities particularly road constructing on accelerating environmental hazardous. It also can be concluded that how multivariate regression model highlights the interrelation existing between independent variables.

The landslide susceptibility map prepared in the frame of the present work is a step forward in the management of landslide hazard in the Tajan river basin located in north of Iran. The LR methodology has demonstrated itself to be a suitable tool when the relationships between landslides and causal factors have to be analyzed. Such a result is achieved by the inspection of the regression

coefficients that determine the role played by influencing factors on the investigated phenomenon. Up to 3.2% of the whole area was assigned to the “very high” susceptibility level. This reveals that the areas have no much prone to landslide. Some weaknesses of this methodology can be pointed out as follows. Firstly, the analysis is still based on an input-output system due to the lack of full statistical capacity within the main GIS packages. In applying the LR model to the geographical data, an external package was necessary for the statistical analysis. However, these packages do not include advanced tools supporting the final mapping of results produced by the analysis and so the resulting data have to be reintroduced into the GIS environment. Secondly, owing to the low scale data used for such regional studies, the results are not very useful on a site-specific scale, where more detailed information and the geo-mechanical properties of landslides have to be considered.

REFERENCES

- 1- Ayalew, L. and Yamagishi, H. (2005): The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda- Yahiko Mountains, Central Japan, *Geomorphology*, 65(1–2), 15–31. Available in [http://www.h5.dion.ne.jp/~hiromitu/5\)Application.pdf](http://www.h5.dion.ne.jp/~hiromitu/5)Application.pdf)
- 2- Constantin Mihaela, Martin Bednarik, Marta C Jurchescu and Marius Vlaicu (2011): Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania); *Environ. Earth Sci.* 63(2) 397–406. Available in <http://link.springer.com/article/10.1007%2Fs12665-010-0724-y>
- 3- Chau, K. T. and Chan, J. E. (2005): Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island, *Landslide*, 2, 280–290, 2005. Available in <http://link.springer.com/article/10.1007%2Fs10346-005-0024-x>
- 4- Chen, Z. and Wang, J.(2007): Landslide hazard mapping using logistic regression model in Mackenzie Valley, Canada, *Nat. Hazards*, 42(1), 75–89. Available in <http://link.springer.com/article/10.1007%2Fs11069-006-9061-6>
- 5- Duman, T., Y., Can, T., Gokceoglu, C., Nefeslioglu, H. and Sonmez, H., 2006. Application of logistic regression for landslide susceptibility zoning of Cekmece Area, Istanbul, Turkey, *Environment Geology*, 51, 241-256. Available in <http://link.springer.com/article/10.1007%2Fs00254-006-0322-1>
- 6- Fenghau Su, Peng Cui, Jianqiang Zhang and Lingzhi Xiang (2010): Susceptibility assessment of landslides caused by the Wenchuan earthquake using a logistic regression model; *J. Mountain Sci.* 7(3) 234–245. Available in <http://link.springer.com/article/10.1007%2Fs11629-010-2015-1>
- 7- Ercanoglu and Temiz (2011): Application of logistic regression and fuzzy operators to landslide susceptibility assessment in Azdavay (Kastamonu, Turkey). *Environ Earth Sci* (2011) 64:949–964, DOI 10.1007/s12665-011-0912-4. Available in <http://link.springer.com/article/10.1007%2Fs12665-011-0912-4>
- 8- GORSEVSKI, P.V.; GESSLER, P.E.; BOLL, J.; ELLIOT, W.E.; FOLTZ, R.B., (2006): Spatially and temporally distributed modeling of landslide susceptibility. *Geomorphology* 80: 178–198. Available in <http://www.sciencedirect.com/science/article/pii/S0169555X06000821>
- 9- Komac M (2006) A landslide susceptibility model using the Analytical Hierarchy Process method and multivariate statistics in perialpine Slovenia. *Geomorphology* 74:17–28. Available in: <http://www.sciencedirect.com/science/article/pii/S0169555X05002072>
- 10- Keefer, D. K. and Larsen (2007), M. C.: Assessing Landslide Hazards, *Science*, 316, 1136–1138. Available in <http://www.sciencemag.org/content/316/5828/1136>
- 11- Kanungo, D. P, Sarkar S., Shaifaly Sharma (2011): Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides; *Nat. Hazards* 59(3) 1491–1512. Available in <http://link.springer.com/article/10.1007%2Fs11069-011-9847-z>

- 12- Khazar.C.C, (1985), Flood mitigation project in Tajan area, final report, Mazandaran,Iran, pp.152-153. Not available in web
- 13- Knapen, A., Kitutu, M. G., Poesen, J., Breugelmans, W., Deckers, J. and Muwanga, A., (2006): Landslides in a densely populated county at the footslopes of Mount Elgon (Uganda): Characteristics and causal factors. *Geomorphology*, 73, 149–165. j.geomorph.2005.07.004. Available in <https://lirias.kuleuven.be/handle/123456789/76875>
- 14- Lan H X, Zhou C H, Wang L J, Zhang H Y and Li R H (2004): Landslide hazard spatial analysis and prediction using GIS in the Xiaojiang watershed, Yunnan, China; *Eng. Geol.* 76 109–128. Available in <http://www.sciencedirect.com/science/article/pii/S0013795204001310>
- 15- Lee, S., Choi J, Min K (2004), Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea. *Int J Remote Sens* 25:2037–2052. Available in <http://www.tandfonline.com/doi/abs/10.1080/01431160310001618734#.UsTyEvQW1qU>
- 16- Lee, S., 2005. Application and cross-validation of spatial logistic multiple regression for landslide susceptibility analysis. *Geoscience*, 9, 63-71. DOI: 10.1007/BF02910555. Available in <http://link.springer.com/article/10.1007%2FBF02910555>
- 17- Lee, S. and Sambath, T. (2006): Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models, *Environ. Geol.*, 50(6), 847–855. Available in <http://link.springer.com/article/10.1007%2Fs00254-006-0256-7>
- 18- Lee, S. and Pradhan, B.(2007): Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models, *Landslides*, 4, 33–41. Available in <http://link.springer.com/article/10.1007%2Fs10346-006-0047-y>
- 19- McCullagh, P., Nelder, J.A. (1989). *Generalized Linear Models (Monographs on Statistics and Applied Probability)*. Chapman & Hall, London, Second Edition , 532 pages. Available in http://www.goodreads.com/book/show/813801.Generalized_Linear_Models
- 20- Nefeslioglu et al. (2008), Landslide susceptibility mapping for a part of tectonic Kelkit Valley. *Geomorphology* 94 (2008) 401–418. Available in <http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.elsevier-437d82e4-868c-3945-b605-6c88f127042a>
- 21- Pradhan (2010): Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling, *Environ. Modell. Softw.*, 25(6), 747–759. Available in <http://www.sciencedirect.com/science/article/pii/S1364815209002886>
- 22- RICKLI, C. & GRAF, F. 2009. Effects of forests on shallow landslides – case studies in Switzerland. *Forest Snow and Landscape Research*, vol.82, pp.33–44.
- 23- REGMI, N. R., GIARDINO, J. R. & VITEK, J. D. 2010. Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. *Geomorphology*, 115 pp.172-187.
- 24- SUJATHA, E. R., RAJAMANICKAM, G. V. & KUMARAVEL, P. 2012. Landslide susceptibility analysis using Probabilistic Certainty Factor Approach: A case study on Tevankarai stream watershed, India. *Journal of Earth System Science*, vol.121,, October 2012,, pp. 1337–1350.
- 25- Sützen ML, Doyuran V (2004) Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. *Eng Geol* 71:303–321. Available in : <http://www.sciencedirect.com/science/article/pii/S0013795203001431>
- 26- van Westen CJ, van Asch TWJ, Soeters R (2006) Landslide hazard and risk zonation: why is it still so difficult? *Bull Eng Geol Environ* 65(5):167–184.available in: <http://link.springer.com/article/10.1007%2Fs10064-005-0023-0>

-
- 27- Vijith, H. and Madhu, G.(2008): Estimating potential landslide sites of an upland sub-watershed in Western Ghat's of Kerala (India) through frequency ratio and GIS, *Environ. Geol.*, 55, 1397–1405. Available in <http://link.springer.com/article/10.1007%2Fs00254-007-1090-2>
- 28- Yalcin, A., 2005. An investigation on Ardesen (Rize) region on the basis of landslide susceptibility. PhD Thesis, Karadeniz Technical University, Trabzon, Turkey. Available in www.liebertonline.com/doi/abs/10.1089/ees.2006.0158.
- 29- Yalcin A (2008) GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Anderson (Turkey): comparison of results and confirmations. *Catena* 1:1–12 available in: <http://www.sciencedirect.com/science/article/pii/S0341816207000173#>
- 30- Yilmaz I, 2010. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: Conditional probability, logistic regression, artificial neural networks, and support vector machine; *Environ. Earth Sci.*, doi: 10.1007/s12665-009-0394-9. Available in <http://link.springer.com/article/10.1007%2Fs12665-009-0394-9>
- 31- Zhu Lei and Huang Jing-Feng (2006): GIS-based logistic regression method for landslide susceptibility mapping on regional scale; *Journal of Zhejiang University Science A* 7(12) 2001–2017. Available in <http://link.springer.com/article/10.1631%2Fjzus.2006.A2007>
-

