J. Basic. Appl. Sci. Res., 3(3)924-930, 2013
© 2013, TextRoad Publication

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Landslide Susceptibility Mapping in Central Zab Basin in GIS-Based Models, Northwest of Iran

Saeed Khezri*[1]; Himan Shahabi[2]; Baharin Bin Ahmad[2]

[1] Department of Physical Geography, Faculty of Natural resources, University of Kurdistan, Iran
[2] Department of Remote sensing, Faculty of Geo Information and Real estate, Universiti Teknologi Malaysia (UTM)

ABSTRACT

There are several practical methods in landslide susceptibility of which the logistic regression is used as the statistical model in central Zab basin in the southwest mountainsides of West-Azerbaijan province in Iran to predict landslide susceptibility with two independent and dependant variables. This part of Zab basin is landslide-prone given its geological structure and geomorphology. We studied and defined the factors (slope, aspect, elevation, distance to road, distance to drainage network, and distance to fault, land use, precipitation, and geological factors) that affect occurrence of the landslides. To get more precision, speed and facility in our analysis, all descriptive and spatial information was entered into GIS system. The applied statistical approach is appropriate to landslide prediction. It employs the landslide events as dependant variable and data layers as independent variable, and makes use of the correlation between these two factors in landslide susceptibility. Given the employed model and the variables, signification tests were implemented on each independent variable, and the degree of fitness of susceptibility mapping was estimated; finally the map was classified into five categories: very low, low, moderate, high and very high risk. The categories cover an area of 95.46km^2, 100.46km^2, 46.1km^2, 158.38km^2 and 120.96km^2, respectively.

KEYWORDS: Geographic Information System (GIS), logistic regression, independent variable, dependant variable, central Zab basin.

INTRODUCTION

As a natural hazard, landslide causes great damages to human lives and infrastructure [1]. Identifying its different types and formation process counts as a major factor in preventing landslide induced damages. Nowadays, statistical methods are more applicable for prediction and classification of environmental problems in various regions [2]. The first paper dealing with such a statistical approach was published almost 25-years ago, Carrara (1988) later, the author modified his original methodology to the GIS environment [3,4].

As a natural disaster, landslides frequently produce a lot of damages to property and human lives all over the world. The preliminary study in Iran estimates an annual financial damage of over 500 billion Iranian Rials. Based on the published results, till mid-1999 a total of 2590 landslide events recorded in the Iranian territory have accounted for 162 casualties, 176 destroyed houses, and creation of 963807m^3 of sediment grains [5].

During the past few years, quantitative methods have been implemented for landslide susceptibility mapping studies in different regions [6]. Nowadays, statistical and GIS methods include logistic regression are more applicable for prediction and classification landslides in various regions. Landslide susceptibility assessment in a logistic regression is based upon suitable selection of factors which play a dominant role in landslide susceptibility.

Preparing susceptibility maps represents a vital step towards identifying unstable regions of the country. Landslide susceptibility maps are useful for taking appropriate precautionary measures to limit land usage in areas of landslide susceptibility, which in turn will contribute to decreased damages [7].

Of the several available statistical models, the multivariable regression or logistic regression model, which is preferred in the present study, is of particular importance. Central basin of Zab River, with its central portion comprising a tectonic valley being the main focus of the present study, displays various slope movements, in particular landslide. Hence, the aim of this study was to use widely-accepted GIS models, statistical methods include logistic regression to analysis approach and landslide causative factor databases developed using Satellite images and aid with GIS in central Zab basin. The methodical

***Corresponding Author:** Saeed Khezri, Department of Physical Geography, Faculty of Natural resources, University of Kurdistan, Iran. Email: SKhezri@uok.ac.ir

procedure in preliminary geological investigation stages presents low cost research, especially for larger areas and lined structures which are endangered both by extremely slow landslides and by rapid debris flows. The scientific contributions of this study are:

1) Identification of the best way to calculate landslide probabilities based on the characteristics of the landslide inventory.

2) Evaluation of the degree of logistic regression model fit by statistical tests and comparisons with the landslide inventory.

3) Evaluation of the best indirect susceptibility map in comparison with a direct susceptibility map

Geographical location of study area

Zab basin occupies southwestern section of West Azerbaijan and northwestern part of Kurdistan. The area under present study covers parts of mountains and slopes in southwestern West Azerbaijan in the central portion of Zab basin between the latitudes of (36° 8' 25") N and (36° 26' 27") N and the longitudes of (45° 21' 21") E and (45° 40' 44") E.

Central Zab basin has a north-south orientation and stretches almost 30 km in east-west direction. The study area covers some 520km^2 of its total area (Fig 1). It is one of the settled geographical basins including a city, three towns or small cities, and over 80 villages [8]. Here, a north-west extension branches off from the east-west oriented ridges of Zab valley, creating a different landscape from that of the internal sections of Azerbaijan and Kurdistan.

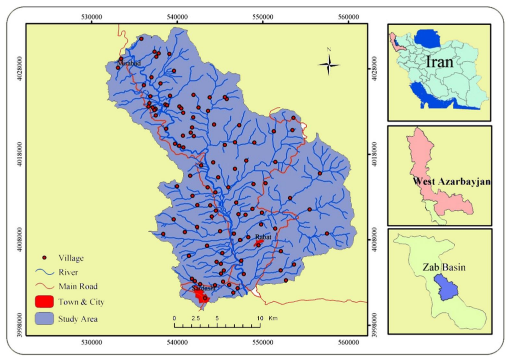

Fig 1: Geographical location of study area

MATERIAL AND METHODS

Materials

In landslide susceptibility mapping, the instable regional factors that their fluctuations were accompanied by differing frequencies of landslide events were defined as controlling factors in susceptibility. They include geology (lithology), geomorphology (elevation, slope, and aspect), distance to roads, distance to fault, land use, precipitation, and distance to drainage network.

First of all, a digital elevation model (DEM) was generated from a triangulated irregular network (TIN) model that was derived from digitized contours with a contour interval of 25 m. The elevation, slope angle, and aspect, data were obtained from the DEM.

The critical point was the selection of appropriate pixel size for positional accuracy and precision of susceptibility levels in the resultant map. The positional accuracy needed for 1:50000 scale maps must be 150 m. For this reason, a pixel size of 50 m was selected for our DEM. Fault lines were derived from 1:100000 scale geology maps and the aerial orthogonal distance of all pixels to fault lines calculated. A similar process was carried out for road lines, and drainage networks. In addition, the kilometer square density of drainage networks, road and fault lines were also used to demonstrate the importance of the features in the whole study area.

The properties of the landslides were recorded on a standard landslide inventory data sheet, but the main purpose herein was to map only the boundaries of the landslides. A digitized map of landslide boundaries was produced, and these digitized maps were input into GIS. A vector-to-raster conversion was performed to provide a raster data of the landslide areas.

Waterways of zab basin were digitized and all the needed operations for using this information in GIS software were carried out. The rasterized pixel size is considered to be 25×25 square meter. We drew the annual co-rainfall curves in Zab river basin by using 30 year statistics of the stations in the studied region, Interpolation technique and Krigging in GIS environment.

Another dataset was land use, which was interpreted from Landsat ETM$^+$ image on the 21 April 2009. Landslide-inventory map of the study area was identified by SPOT 5 satellite on the 25 May 2008 extensive field studies were used to check the size and shape of landslides, to identify the type of movements and the materials involved, and to determine the state of activity (active, dormant, etc.) of the landslides.

Method

Logistic Regression

Logistic regression, which is one of the multivariate analysis models, is helpful for forecasting the presence or absence of a characteristic or outcome based on the values of a set of predictor variables. The advantage of logistic regression is that, through the addition of a suitable 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 [9].

The coefficients were estimated using the maximum likelihood model. In other words, the coefficients that made the observed results most "likely" were selected. Since the relationship between the independent variables and the probability is nonlinear in the logistic multiple regression model, an iterative algorithm is required to estimate the parameters [10].

RESULTS AND CONCLUSION

Slide zones layer was defined as the dependant variable. Data on this layer was first acquired as points using GPS. Then, slide zones were extracted using these points and satellite images. Finally, as a raster layer, the landslide layer was extracted as a binary dependant variable with 0 (absence of landslide) or 1 (presence of landslide) values (Fig 2).

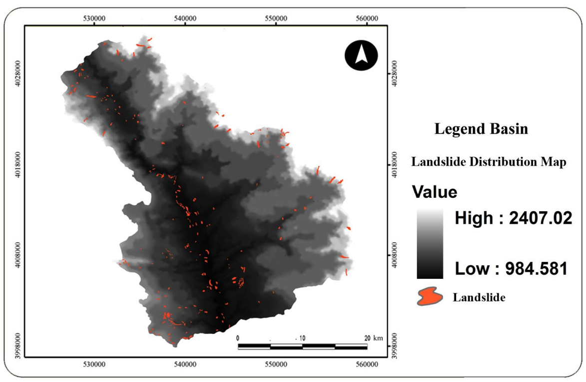

Fig 2: Landslide distribution map of central Zab Basin

Geology

Geological factor is considered as independent variable in landslide formation in that lithology and its varied structure tends to lead to a variation in stone stability and strength and also to a varied soil texture [10]. The layer was extracted from 1:100000 maps of the Iranian Geology Organization. It was quantified and then appended to the model since it proved to be statistically significant. Most of slide events in the study area occurred in loose formations including fine-grained sediments in particular in alluvial terraces. Stones, homogenous phylite formations, marble, lime, green andesite, dolomite and sandstone have the widest distribution in the region. Proving statistically significant, the geological layer was classified in terms of stone stiffness (Table 1).

Table 1: Classification of geological strata

Lithology	Value
Sandstone, andesite	1
Lime (dolomite-orbitolina)	2
Lime and shale and sand	3
Phylite and gneiss	4
Shale	5
Alluvium	6

Elevation

While landslide occurs in a certain elevation accompanied with a certain gradient, the elevated height of a given region results in an increased gradient which in turn increases the shear forces in slopes. The raster layer of Digital Elevation Model was prepared using 1:50000 maps of the Iranian Geology Organization and their interpolation.

Distance to fault

Fault in itself does not act as a primary factor in landslide formation; higher distribution of faults, however, may contribute to a higher slope gradient which in turn results in increased gravity down the slope [11]. Several faults in the central Zab basin have created faces in different escarpments. As a consequence, discontinuities and interruptions are evident between different contexts.

Slope

Slope is viewed as the major controlling factor in landslide formation. Critical slope intensifies the influence of the other factors. In a uniform slope, with homogenous lithological characteristics, increased slope has the major influence on creation of landslide [12]. While the maximum slope is not necessarily associated with maximum slide, there will be almost no chance of landslide without adequate slope angle.

Slope aspect

Slopes with a westerly aspect typically have excessive snow stability, and this is followed by increased moisture in such slopes. This justifies the selection of this important parameter as a determining factor in landslide formation. In general, northern slopes are influential in creation of landslide due to snow durability and moisture, in particular during spring. Further, these slopes cause tectonic activities as a result of longer moisture durability during warm seasons, which is associated with creation of fine-grained deposits that have the highest susceptibility of landslide [13]. Given the importance of this factor, a map of slope aspect was also prepared for the study area, and was incorporated in regression model.

The layer proved qualitative; therefore, it was classified into three categories: northern; western and eastern; and southern, southwestern and southeastern (Table 2).

Table 2: Classification of slope aspects

Slope aspect	Value
N-NE-NW	3
E-W	2
S-SW-SE	1

Distance to road

Human usages and his activities have always played a decisive role in environmental changes. The site selection for these usages occasionally is such inappropriate that leads to disorganized and interrupted natural ecosystem. An instance of such inappropriate site selection concerns road construction in which any negligence during its prospective studies will amount to endangering human lives [14]. In sum, there is a significant correlation between the extent and frequency of landslide on one hand and distance to regional road system on the other. The parameter causes cuts in slopes and thereby destabilizing them.

Density of drainage system

Drainage layer was selected so as to consider the roles of inappropriate drainage and undercutting of hydrographic system, simultaneously. Inappropriate drainage systems promote infiltration of water into slope. Higher number and density of drainage systems running on or along the base of slope will bring about more desirable situation for water infiltration. Density of drainage system is among the controlling factors of landslide formation. In studied area, localities with decreased slope gradients revealed increased density of watercourses. This has significantly affected creation of landslides. In this study, the assumed radius for estimating the drainage system density was 100 km. Very low density of hydrographic network the studied area is indicative of intense infiltration of surface runoff into the textures of central Zab basin, which provides conditions for the phenomenon under consideration.

Precipitation

Precipitation, in particular sudden, intense rains, and snow melting is a controlling factor which trigger mass movements through providing water, increasing the underground hydrostatic level and pore water pressure. Since the studied area is second only to the northern regions in Iranian territory in terms of precipitation rate, the regional precipitation distribution pattern was mapped and incorporated into the model.

Land use

Changes in ground surface have the potential to affect slope failure, given the presence of other relevant factors. The lack of appropriate coverage such as vegetation is influential in increased frequency of landslides. In addition, loading the slope, and in particular digging the base of slope and undercutting and injecting water into it are effective elements in landslide occurrence [15]. Thus, plotting the slide layer against that of the land use reveals that the frequency and density of landslides in central Zab basin is higher in arid landscapes.

This qualitative variable was also categorized by the relative importance of each usage (Table 3).

Table 3: Classification of land use with relative value of individual usages

Kind of land sue	Natural forest	First-class pasture	Second-class pasture	Man-made forest	Dry farm land	Barren land	Settled
Value	1	2	3	1	4	5	5

Application of Logistic regression model in Landslide susceptibility mapping

Here, data-based approach was used in landslide susceptibility map. The method is based on using available data on landslide events. The practical statistical methods in data-based approaches to landslide susceptibility are regression model. The model uses the available data to predict the independent variable (Y) from one or more independent predictive variables (X1) by a certain relation. However, if the independent variable Y is qualitative and always has a binary value (occurrence [1] nonoccurrence [0]), the ordinary regression will fail to establish the relation [16]. In such cases, the probability of phenomenon (P) is estimated, and the alternative model employed is logistic regression analysis, given by:

$$Y = \mathrm{Logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \tag{1}$$

$$\mathrm{Logit}(p) = \mathrm{Ln}\left(\frac{p}{1-p}\right)$$

Where, P is occurrence probability, β0 is the constant, X1 to Xn are the independent variables, and β1 to βn are the coefficient of these variables, respectively. In effect, Eq. 1 estimates the Logit (p) values and extracts the p from these values. The above relation can be normalized using the least squares; this procedure, however, has limitations since the P values are not given.

In logistic regression modeling, maximizing likelihood function is used to obtain the coefficients of independent variables. Likelihood function is in fact a probability. In particular, it is a probability that the observed values of dependant variable might be predicted by those of independent variables. Likelihood function is defined as Eq. 2. In this method, the likelihood function—Eq. 2—is maximized to obtain the coefficients.

$$L = \prod_{i=1}^{N} p_i^{y_i} \times (1-p_i)^{(1-y_i)} \tag{2}$$

Where, pi is the probability of slide occurrence defined by Eq. 1, and yi represents the presence (1) or absence (0) of the observed mass movement. Likelihood ratio index is used if evaluation of relative significance of any individual independent variable is required in the process of predicting the dependant variables through logistic regression model. The index equals to -2 Log (L), and will have a value ranging between 0 to +∞, define by the L range. Significance of an independent variable will be higher than that of another independent variable when its exclusion results in a wider variation in the likelihood ratio. The linear model provided by logistic regression analysis rests on slides occurred with respect to independent variables or earlier situations that led to landslide [17]. Therefore, to obtain the optimized regression coefficient landslide layer was defined as the response variable (Y) and mass movement presence was designated as 1 and its absence as 0. Other extracted layers (distance to road, gradient, elevation, slope aspect, distance to fault, geology, drainage density, and land usage were defined as the predictive variables. Since the layers of slope aspect, land usage and geology were qualitative; they were classified based on their significance and influence on slope failure. Finally, the entire layers were standardized using Eq. 3. Some layers, including distance to river, were excluded from equation as not being qualitative.

$$X_{S\tan dard} = \frac{X - Min(x)}{Max(x) - Min(x)} \tag{3}$$

Once the effective layers in susceptibility were identified, the qualitative layers were classified and then were standardized by Eq. 3 using GIS. Thus, the value of all independent variables (X) ranged from 0 to 1. All independent variables are equal to 1, and if no slide is present the value is 0. Then, values of dependant and independent variables were sampled randomly, and in SPPS software the two variables were correlated to extract their relative significance and the coefficients of each independent variable.

The following tables summarize the results from correlation of independent and dependant variables. Independent variables with weak correlation have been excluded from the tables. As it is shown, the variables have a sig. or significance value of 0. If the value is lower than 0.5, the correlation between the dependant and independent variables are considered significant (Tables 4-5). Thus, the model was applied to extract the coefficients of the whole variables. The Change column in Table 5 shows the significance of individual variables. Thus, distance to fault variable has the highest significance with negative correlation as reflected in column B in Table 4. In other words, the increased distance to fault will result in decreased landslide frequency. Distance to road variable takes the second place, while lithology layer represents the least significant variable. Frequency of landslides correlated positively with the type of stones, thus the highest frequency was recorded in areas with less stiffer stones.

Table 4: Significance coefficient of variables

Variables in the Equation						
Variables	B	S.E.	Wald	df	Sig.	Exp(B)
precipitation	0.648	0.042	238.421	1	0.000	1.911
Distance to road	-1.068	0.057	352.394	1	0.000	0.344
Drainage density	1.197	0.094	162.378	1	0.000	3.311
Distance to fault	-5.137	0.105	2395.961	1	0.000	0.006
Slope	1.075	0.053	411.677	1	0.000	2.931
Slope aspect	0.116	0.014	66.415	1	0.000	1.122
Lithology	0.041	0.007	35.663	1	0.000	1.042
Land use	0.079	0.007	139.874	1	0.000	1.082
constant	-0.590	0.047	158.003	1	0.000	0.555

Table 5: Coefficients of independent variables derived from logistic regression model

Model if Term Removed					
Variable		Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change
Step1	precipitation	-32375.887	240.008	1	0.000
	Distance to road	-32435.035	358.305	1	0.000
	Drainage density	-32338.200	164.634	1	0.000
	Distance to fault	-33879.959	3248.152	1	0.000
	Slope	-32465.230	418.693	1	0.000
	Slope aspect	-32289.140	66.514	1	0.000
	Lithology	-32273.693	35.619	1	0.000
	Land use	-32326.007	140.249	1	0.000

Thus, landslide susceptibility mapping was carried out based on the obtained coefficients. To test the accuracy of the model, slide zones model as points was overlaid on the landslide susceptibility mapping (Fig 3).

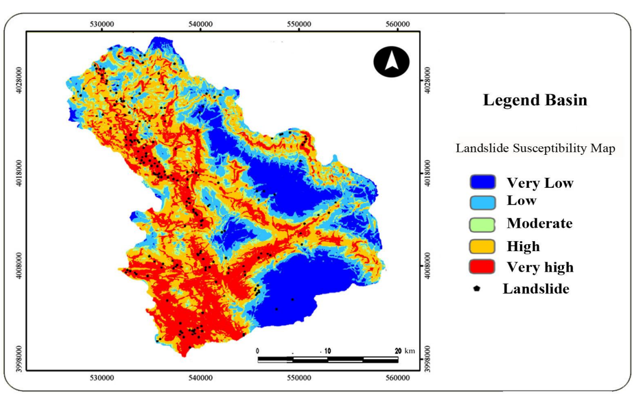

Fig 3: Landslide susceptibility map of central Zab basin

Once classified, area and area percentage of each class were also estimated (Table 6).

Table 6: Landslide risk and risk classes as area and area percentage

Risk class	Very low	low	moderate	high	Very high
Area (km^2)	95.642	100.4672	48.1008	156.3832	120.96
Area %	18.3	19.32	9.4	30	23

Acknowledgements

We are thankful to the Department of Remote sensing in Universiti Teknologi Malaysia and Department of Physical Geography in University of Kurdistan and UTM International Doctorial Fellowship (IDF) for providing the facilities for this investigation.

REFERENCES

[1] Nefeslioglu, H.A, C, Gokceoglu and H, Sonmez, 2008. An Assessment on the Use of Logistic Regression and Artificial Neural Network with different Sampling Strategies for the Preparation Maps of Landslide Susceptibility. Engineering Geology., 97: 78-89.

[2] Paliwal, M, U.A, Kumar, 2009. Neural Networks and Statistical Techniques: A Review of Applications. Expert Systems with Applications., 36: 2-17

[3] Carrara, A, 1988. Landslide hazard mapping by statistical method: a ‘‘black box” approach. In: Proceedings of Workshop on Natural Disaster in European Mediterranean Countries. Conzilio Nazionale delle Richerche, Perugia, Italy: 205–224

[4] Carrara, A., M, Cardinali, R, Detti, F, Guzzetti, V, Pasqui, P, Reichenbach, 1991. GIS techniques and statistical models in evaluating landslide hazard. Earth Surface Processes and Landforms 16: 427–445

[5] Karam, M, 2001. Quantities models and landslide hazard zonation in crashed Zagros-case study: Sarkhon basin in Chahar mahal Bakhtyari county, Ph.D. thesis in Tarbyat Moalem University.

[6]Luzi, L., Pergalani, F and Terlien, M.T.J, 2000. Slope vulnerability to earthquakes at Sub regional scale, using probabilistic techniques and geographic information systems, Engineering Geology, 58, 313–336.

[7] Shahabi, H., B.B Ahmad, S, Khezri, 2012. Evaluation and comparison of bivariate and multivariate statistical methods for landslide susceptibility mapping (case study: Zab basin). Arabian Journal of Geosciences, 1-23.

[8] Khezri, S, S, Roostaei, and A.H Rajaie, 2006. Evaluation and mass movement hazard zonation in central section of Zab basin by using Anabalagan method, Humanities science Journal (issue geography), 38: 32-42.

[9] Lee, S, 2005. Application and cross-validation of spatial logistic multiple regression for landslide susceptibility analysis. Geosciences, 9(1): 63-71.

[10] Ayalew, L and H, Yamagishi, 2005. The Application of GIS-based Logistic Regression for Landslide Susceptibility Mapping in Kakuda-Yahiko Mountains, central Japan. Geomorphology 65: 15-31

[11] Ohlmacher, G.C and J.C, Davis, 2003. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Eng. Geol 69: 331–343

[12] Dai, F.C and S, Lee, 2002. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong, PP. 213-227.

[13]Gruber, S, C, Huggle, and R, Pike, 2009. Modeling Mass Movements and Landslide Susceptibility. Developments in Soil Science 33:76-86

[14] Castellanos, E.A and C.J, Van Westen, 2007. Generation of a landslide risk index map for Cuba using spatial multi-criteria evaluation. Landslides 4: 311–325

[15] Fall, M, R, Azzam and C. Noubactep, 2006. A multi- method approach to study the stability of natural slopes and landslide susceptibility mapping, Engineering Geology,82: 206-207.

[16] John, C , C, Davis and D, Ohlmacher, 2006. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA , Computers & Geosciences, 32: 1122-1123.

[17] Lee, S and T, Sambath, 2006. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. Environ. Geol 50: 847–855.