JOURNAL GEOLOGICAL SOCIETY OF INDIA
Vol.84, July 2014, pp.68-86

Landslide Susceptibility Zonation Mapping Using Logistic Regression and its Validation in Hashtchin Region, Northwest of Iran

REZA TALAEI
The Center of Agriculture and Natural Resources of Ardabil, Ardabil, P.O. Box 56135-545, Iran
Email: rztala@yahoo.com

Abstract: Landslide susceptibility zonation mapping assists researchers greatly to understand the spatial distribution of slope failure probability in a region. Being extremely useful in reducing landslide hazards, such maps could simply be produced using both qualitative and quantitative methods. In the present study, a multivariate statistical method called 'logistic regression' was used to assess landslide susceptibility in Hashtchin region, situated in west of Alborz Mountains-northwest of Iran. In this study, two independent variables, categorical (predictor) and continuous, were drawn on together in the model. To identify the region's landslides use was made of aerial photographs, field studies and topographic maps. To prepare the database of factors affecting the region's landslides and to determine landslide zones, geographic information system (GIS) was used. Using such information, landslide susceptibility modeling was accomplished. The data related to factors causing landslides were extracted as independent variables in each cell (in 50 m×50 m cells). Then, the whole data were input into the SPSS, Version 18. The prepared database was later analyzed using logistic regression, the forward stepwise method and based on maximum likelihood estimation. Regression equation was determined using obtained constants and coefficients and the landslide susceptibility of the area in grid-cells (pixels) was computed between 0 and 0.9954. The Receiver Operating Characteristic (ROC) curve was used to assess the accuracy of the logistic regression model. The predicting ability of the model was 84.1% given the area under ROC curve. Finally, the degree of success of landslide susceptibility zonation mapping was estimated to be 79%.

Keywords: Landslides Susceptibility, Logistic Regression, Geographic Information System (GIS), Iran.

INTRODUCTION

Landslides are considered as one of the most widespread and demanding natural hazards causing billions of dollars of damage to property and human life every year (Wang et al. 2006; Ilinca and Gheuca, 2011; Aleotti and Chowdhury, 1999; Dai et al. 2002; Yesilnacar and Topal, 2005). In Iran, most landslides occur in Alborz mountains, in the northeast and northwest and Zagros mountains, in the northwest and southeast parts (Shoaei et al, 2005). Due to its geological, climatic and tectonic activities, Alborz mountains witnesses many landslides each year, which cause damage to life and property (Jadda et al. 2009). Currently, there is no reliable report specifying the amount of damage caused by landslides in Iran, but some informal reports have estimated that only the amount of direct damage mounts to 50 million dollars (Komakpanah and Hafezimoghadasi, 1994). Hashtchin is located in northwest Iran and is situated along the western slope of Alborz mountains and is home for two different types of landslides: (1) New and active landslides, (2) reactivation of old landslides. The exact amount of damage to Hashtchin area by such landslides is not known, yet there is no doubt that such landslides can bring about a large number of direct and indirect damages to people and land (Mahdavifar, 1996; Talaei et al. 2004). In fact, there is an urgent need for a modern developmental and strategic plan to mitigate landslide damage and hazard in hilly regions like Hashtchin. To fulfil this objective, susceptible zones in the area should be determined based on the geological, topographic, geomorphologic features and human activity. This could be accomplished with the help of landslide susceptibility zonation mapping in which a region is divided into several zones, each with a different degree of landslide susceptibility (Anbalagan, 1992). To prepare susceptibility maps developments have been made to estimate spatial variations in slope failure probability of a region (Mathew et al. 2009). Different methods have been proposed to generate landslide susceptibility maps. The type, precision and scale of such maps depend on factors selected and their use (Varnes, 1984; Van Westen, 2000). Therefore, for landslide susceptibility zonation mapping, it is necessary to understand comprehensively factors contributing to landslides in a region and to assess the relative and

0016-7622/2014-84-1-68/$ 1.00 © GEOL. SOC. INDIA

cumulative impact of such factors. From among the methods used to assess landslide susceptibility, heuristics, statistical and deterministic approach are quite well known (Soeters and Van Westen, 1996). Further, during the past few years multivariate statistical approach for the landslide susceptibility assessment has been used (Meusburger and Alewell, 2009). In heuristic or direct approach, expert views are drawn for landslide susceptibility zonation mapping (Niemann and Howes, 1991; Anbalagan, 1992; Turner and Schuster, 1996; Atkinson and Massari, 1998; Van Westen et al. 1999). One major problem with this approach is that it requires abundant geological and environmental information about landslides and factors contributing to its occurrence. This volume of information is, of course, really difficult to obtain, and at times even impossible to access. Other shortcomings of this approach are: reproducibility of its results; subjectivity of the weighting system used, and ranking and grading of the variables (Dai et al. 2001). The deterministic approach is used for large-scale landslide susceptibility assessments. It may be used if two conditions are met: First, the geological as well as geomorphological characteristics of the whole region must be homogenous; Second, the type and the nature of the landslide must be simple, well-known and well identified (Dai et al. 2001; Turner and Schuster, 1996). The main advantage of this approach is that it allows factor of safety to be computed for a slope. In small areas, slope stability cannot be computed with desired accuracy using other methods but the deterministic approach can handle this issue successfully (Van Westen, 1993; Terlien et al. 1995, Wu and Sidle, 1995). In statistical methods, all factors contributing to past landslides are dealt with statistically in the whole region and terrain units (Van Westen, 1993; Naranjo et al. 1994). Using data related to landslide-free zones, the probability of a landslide occurring in the future is predicted quantitatively or semi-quantitatively in this area. To analyze landslide susceptibility zonation, widespread use of bivariate and multivariate statistical methods has been made (Brabb et al. 1972; Yin and Yan, 1988; Carrara et al. 1990; Van Westen, 1993; Van Westen et al. 1999; Naranjo et al. 1994; Dai and Lee, 2002; Donati and Turrini, 2002; Ohlmacher and Davis, 2003; Yesilnacar and Topal, 2005; Davis et al. 2006; Pradhan et al. 2008; Mathew et al. 2009; Dwi Wahono, 2010). Due to the suitability of the application of GIS in statistical methods, statistical-based approaches have proved most practical in landslide susceptibility zonation analyses at medium and regional scales (Turnern and Schuster, 1996). Within the multivariate statistical methods for landslide susceptibility zonation analysis, multiple linear regression, discriminant analyses and logistic regression are the most frequently used (Mathew et al. 2009). The present article primarily intends to assess landslide susceptibility zonation in Hashtchin region, located west of Alborz mountains in northwest Iran using multiple logistic regression method.

A number of researchers have carried out studies on landslides and factors causing them in Hashtchin area (Ansari and Blurchi, 1996; Nikandish and Mir Sanei, 1996; Talaei et al. 2004). Hashemi Tabatabaei (1998) studied factors contributing to the occurrence of landslides in the region and produced a regional hazard map using qualitative model. Mahdavifar (1996), as well as Uromeihy and Mahdavifar (2000), studied factors causing landsides in Khoreshrostam area (part of Hashtchin region). In their studies, the Hazard Potential Index (HPI) was calculated by a computer program using fuzzy sets. In fact, the prepared hazard zonation map was a susceptibility map the accuracy of which was not checked or assessed by the authors.

The present study carried out on landslide susceptibility zonation in Hashtchin area and Alborz mountains using logistic regression model is first of its kind. Further, to assess the results of logistic regression method using Receiver Operating Characteristic curve analysis (ROC) and based on the group of selected landslides is not included in the model. The important significance of the present study is that the modeled landslide susceptibility zonation map can be used as a reliable input to determine landslide type along with ways to mitigate its damages. The results can also be used as the basis to design cautionary developmental programs as well as environmental planning and to prevent landslides. The model proposed could also apply to other regions with similar geological, topographic and climatic features as that of the Hashtchin region.

STUDY AREA

Hashtchin region is situated in southwest of Ardabil province in northwest of Iran. The area under study lies between longitudes 48°14' to 48°44'E and latitudes 37° 06' to 37°32'N (Fig. 1). It covers parts of Talesh mountains, Agh Dagh massif, Darram hills, Qezel Owzan valley and gorges in northwest of Alborz mountains. The study area is 1645.84 km^2, and 9.52% of the area is found to be affected by landslides. Currently, more than 9.52% of Hashtchin's territory is struck by isolated and regional landslides (Mahdavifar, 1996; Talaei et al. 2004). A total of 175 landslides (20.89 km^2 single landslides and 135.86 km^2 landslide zones) were mapped in the region covering 156 km^2.

Landslides are classified as translational, rotational slides and combinations of the two. Landslide zones are also

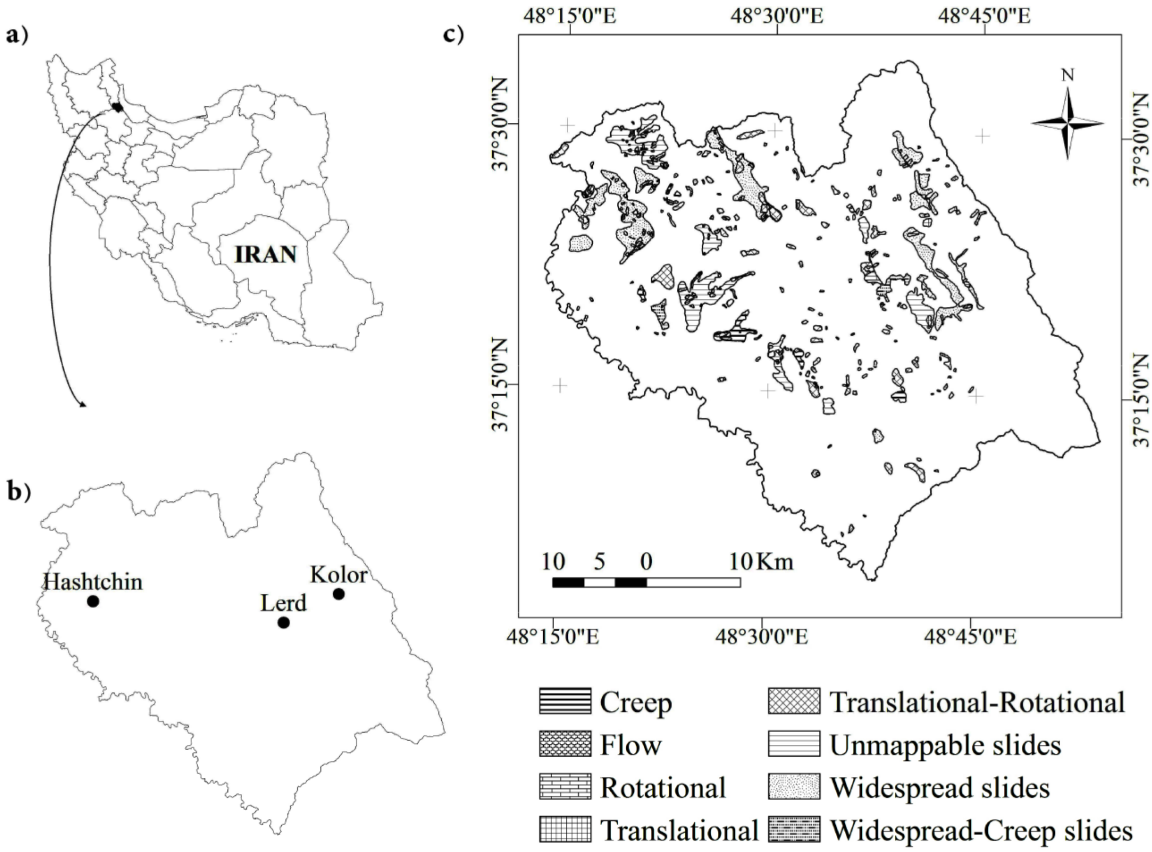

Fig.1. Location map of study area (**a** and **b**) and landslide inventory map (**c**)

classified as creep, unmappable and widespread. It was found that 103 cases (58.9%) of the landslides of the region are presently active. More than 60% of landslides which have been studied show signs of activities in the past 50 years.

METHODOLOGY

For the present study, the following steps were taken: First, the landslide records were entered to standard forms as well as its inventory map was prepared by field work and aerial photographs (scales=1:50000, 1:20000). Data extracted regarding causes of previous landslides can be used as source for predicting future landslides (Wang and Sassa, 2005). For this reason, as the next step, information regarding causes of landslides was collected using ArcGIS 9.2 package. Digital Elevation Model (DEM) was also prepared using topographic maps (scale=1:25000) with a resolution of 10×10 meters. Based on the Digital Elevation Model (DEM) obtained, geomorphologic parameters including slope gradient, slope aspect, profile curvatures and altitude were prepared in GIS system. Due to the importance of lithological and tectonic features of the region in causing landslides, steps were taken to collect the required data with great care. The main mapping tools used included topographic maps at 1:50000 and 1:250000 scales, and aerial photographs at 1:20000 scale and geological maps of Hashtchin (Faridi and Anvari, 1996), Masuleh (Davies et al. 1972), and Bandar Anzali (Davies et al. 1975) at 1:100000 scale. The information already present in the previous maps was used to prepare the geological map. During the field work, geological observations, important lithostratigraphic and geological structures were identified and transferred to the topographic maps. From the region's geological maps, lithological layers and distances to major faults were produced. The data related to the type of plantation (land cover) and land use were extracted from topographic maps and satellite images of Landsat ETM$^+$-2002. Atmosphere and geometric corrections were made for images before processing further. The land cover-land use thematic map of the study area was prepared using unsupervised classification of Landsat ETM+ (2002) satellite images because of the available information on the study area was not sufficient (Bakker et al. 2004). This map was revised and completed based on the observations made during the field works. The major land cover-land use classes in the area are rangeland (> 92%), agricultural land and orchard (5.2%), forest (2%) and built-up area. Vegetation provides both hydrological and mechanical effects that are generally advantageous for the stability of the slopes. Comparing aerial photographs of the years 1958, 1968 and 1993 with 2002 satellite images shows that in this period, many changes have taken place in land cover of the area. For more than 60 years the forest and rangelands have been extensively damaged due to unscientific actions and applications that have been changed into agricultural lands, orchards, and built-up areas to construct roads. It is very difficult to find an area that is unaffected by these changes. Therefore, the land cover-land use map was produced in detail because it has been recognized as one of the important independent variables to be used in the landslide susceptibility analysis.

Earthquake is considered as one of the factors in triggering of landslides. The extent of the impact of earthquake on slope stability depends on geological, lithological, hydrologic, topographic and other related conditions of the region. To simplify the impact of earthquake on landslide, increase shear stresses along a failure surface is considered as the only force contributing to slope instability. To carry out pseudo static method, this force is deemed to be the result of Peak Ground Acceleration (PGA) (Shariat Jafari, 1996). To study the impact of the earthquake on occurrence of landslides, the role played by active faults were studied and the PGA map was also prepared. This map can be used as an important data layer in analyses related to landslide hazard and susceptibility (Turner and Schuster, 1996). Changes in slopes during road construction can reduce soil and layer stability and hence cause landslides (Knapen et al. 2006; Ayalew and Yamagishi, 2005). The layers related to distance to roads and settlements were produced using topographic maps in order to incorporate variables related to roads and buildings in the landslide susceptibility zonation analysis model. The impact of hydrology and climate on occurrence of landslides in the region was assessed in terms of mean annual precipitations and distance from rivers. Eq. 1 was used to determine the mean annual precipitations. This multivariate regression equation was obtained based on meteorological data from 58 meteorological stations, which determines the mean annual precipitation at different points of the region using their geographic coordinates and elevations. Further, the isohyetal map was also prepared on 1:50000 scale (Hemmati et al. 2007).

$$\textit{Mean annual precipitations} = a -(X\times 4.81) +(Y\times 0.002) + (Z\times 0.063) \tag{1}$$

In equation 1, a (is a constant) = 471.162, X=longitude in degrees; Y=latitude in degrees, and Z=altitude (elevation) in meters.

The data related to the independent and dependent variables were saved using nominal and scale group measures as an ASCII file in the form of 50×50 m cells, which were later imported to SPSS Version 18 for statistical analysis.

Different multivariate statistical analyses, i.e. discriminant analysis, multiple regression, and logistic regression, are available to assess landslide susceptibility in widespread and complex areas (Carrara et al. 1991; Carrara, 1983; Lee, 2005; Guzzetti et al., 2006; Chang et al. 2007). Type and measurement level of independent and dependent variables determine which model fits a given situation best (Pallant, 2007).

In multiple regression approaches, the dependent variable is measured as a continuous variable. Such approaches are not suitable when categorical dependent variables are available. In the present work, logistic as well as discriminant analyses seem suitable since, the dependent variable is dichotomous – is binary in nature (landslide/no landslide). In discriminant analysis, the independent variables should have normal distribution and there should not be a linear correlation between them. When the above assumptions are not met, logistic regression model is more appropriate (Pohar et al. 2004). Similarly, when the independent variables are categorical, continuous or a combination of the two, logistic analysis is preferable to discriminant analysis (Atkinson and Massari; 1998 Menard, 2002). In logistic method, the main objective is to find a function that can best show the relationship between landslide/no landslide situations (dependent variable) and a collection of independent variables (causes of landslides) (Ayalew et al. 2005; Kleinbaum and Klein, 2010). Logistic method has been used to predict landslide/no landslide situations based on an array of contributing factors (independent variables). Since multiple logistic regression is a type of extended linear model, its dependent variable is expressed by a digit between 0 and 1. Similarly, the probability of landslide occurring is also expressed using real numbers 0 to 1. Therefore, multiple logistic regression can be a very appropriate mean to analyze landslide occurrence (Yesilnacar and Topal, 2005). Recently several studies have been carried out on the assessment of landslides by logistic regression (Carrara et al. 1991; Atkinson and Massari, 1998; Guzzetti et al. 1999; Dai et al. 2001; Lee and Min, 2001; Dai and Lee, 2002; Ohlmacher and Davis, 2003; Lee, 2004; Ayalew and Yamagishi, 2005; Ayalew et al. 2005; Can et al. 2005; Wang and Sassa, 2005; Yesilnacar and Topal, 2005; Duman et al. 2006; Chang et al. 2007; Greco et al. 2007; Mathew et al. 2009; Dong et al. 2011; Grozavu et al. 2012). Since in the present study, the dependent variable is binary (landslide/ no landslide) and the independent variables are continuous and qualitative, the probability of landslide occurrence in each grid cell was computed using logistic regression method. For categorical variables, dummy variables were used.

The Logistic Regression Model

Logistic regression is a mathematical modeling approach which can be used for the evaluation of the relationship between various independent variables and a categorical outcome (Kleinbaum and Klein, 2010). Conceptually, logistic regression is similar to linear regression model since in both the methods the relationship between a set of independent or predictor variables and a dependent variable is evaluated. However in linear regression, the dependent variable usually has continuous values while in logistic regression it is a dichotomy and the independent variables could be of any type. In the present analysis, the dependent variable is a binary (dichotomous) representing the presence or absence of the landslides and the independent variables (predictors) can be either continuous or discrete. Also it is not necessary for the variables to have a normal distribution. Using the logistic regression model the relationship between the logistic function f(z) and the probability of a landslide occurrence can be defined as (Rupert et al. 2008; Meusburger and Alewell, 2009; Mathew et al. 2009; Dwi Wahono, 2010; Grozavu et al. 2012):

$$f(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}} \tag{2}$$

or

$$z = \log\left(\frac{f(z)}{1-f(z)}\right) \tag{3}$$

Where z varies from -8 to +8 while the range of the logistic function f(z) is between 0 and 1.

To obtain the logistic model from the logistic function, z is written as a linear combination of independent variables and their respective coefficients. Mathematically, the regression model is represented by the following equation (Eq. 4)

$$z = b_0 + b_1 X_1 + b_2 X_2 + ... + b_p X_p \tag{4}$$

Where, b_0: is the intercept (often labeled the constant), $b_1, b_2, ... b_p$ are the coefficients that measure the contribution of independent factors ($X_1, X_2, ... X_p$) to the variation in z. If z was observable, we would simply fit a linear regression to z. However, since z is unobserved, the independent variables must be related to the probability of interest by substituting for z. Thus, z is an index that combines independent variables (e.g. lithology). By substituting Eq. 4 into Eq. 2, we get:

$$f(z) = \frac{1}{1+e^{-(b_0+b_1X_1+b_2X_2+....+b_pX_p)}} \tag{5}$$

The logistic model for landslide occurrence can then be represented as:

$$P(D = 1/X_1, X_2, ..., X_p) = \frac{1}{1+e^{-(b_0+b_1X_1+b_2X_2+....+b_pX_p)}} \tag{6}$$

Where, $P(D = 1/X_1, X_2, ... X_p)$ is the probability of a cell undergoing slope failure, given the presence of independent variables X_1 to X_p. The terms b_0 and b_p in this model represent unknown parameters that are estimated based on the data of independent variables and landslide condition of the cells using the maximum likelihood method. The maximum likelihood is derived from the probability distribution of the dependent variable.

Because of the non-linear relationship between independent variables and the probability in the logistic model, an iterative algorithm is required to estimate the parameters (Dai and Lee, 2002). Logistic regression tries to estimate b_0 and b_p by best fitting the observation variables X_i s for the sample locations for which the status of dependent variable is either present or absent (Mathew et al. 2009). Logistic regression modeling is an appropriate approach to estimate the probability of landslide occurrence and slope instability assessment on a regional scale because by using this model we can represent the presence of a landslide at any given cell with a value of 1 and the absence of a landslide with a value of 0. The logit of f(z) are calculated at each cell using values of dependent and independent variables and the estimated coefficients (b_0 and b_p). The obtained values for f(z) describe the probability of occurrence of a landslide which ranges between 0 and 1. Using these formulas, a landslide susceptibility map was generated.

Tests of Model Fit and Accuracy

In the landslide susceptibility map prepared using the logistic regression method, the probability of a landslide occurrence is shown in each cell within the range 0 to 1. The practical application of the landslide susceptibility map produced requires map classification systems. Four common classification methods are available including grading based on quantiles, natural breaks, equal intervals and finally standard deviations. In the quantile method, for a given class or range of susceptibility, a wide range of probability quantities is considered. When severe fluctuation is observed in the data, the natural breaks method is considered as the best to determine the groups. Of course, this latter method is not appropriate for the probability maps designed. Classification of landslide susceptibility based on probability ranges at equal intervals has been frequently used in most studies. In this method, the relative susceptibility of each class compared to other classes is emphasized. Of course, this method is not recommended since it does not show the real susceptibility level involved. The fourth method is based on standard deviation scores. In this method, half a standard deviation is added or deducted from the average probability quantities to determine the fit class. To determine the other classes, a standard deviation is added or deducted from the quantities obtained in the first class (Ayalew and Yamagishi 2005). The standard deviation method is a good method to determine four susceptibility classes including very low, low, medium and high (Ayalew and Yamagishi, 2005; Dwi Wahono, 2010). In this article, in addition to the application of the standard deviation method, landslide susceptibility quantities were divided into four equal groups based on four degrees. The results of both methods were evaluated. To achieve an accurate assessment of the conformity of the classified map based on the aforementioned method the following equation has been used (Fernandez, 2003).

$$\text{D.F.} = \frac{z_i / s_i}{\sum z_i / s_i} \tag{7}$$

Where D.F. is degree of fit; z_i is the area occupied by the rupture zones in the *i* class of susceptibility and s_i is the area of the *i* class of susceptibility.

The smaller the degree of fit in the very low and low landslide susceptibility classes (relative error), and high conformity in medium and high landslide susceptibility classes (the amount of relative success) represent the high quality of the susceptibility map. Zonation maps can further be assessed by determining the percentage of failure zones in each susceptibility class. This method also enables estimation of absolute errors and relative success (in order of failure percentages in classes with very low, low, medium and high susceptibility ranges).

The goodness of fit statistical procedure assesses the conformity of the logistic model to the real result. The Hosmer-Lemeshow statistics is one of the goodness of fit statistics that shows the conformity of observed cases to the expected ones for two member classes (in this article landslide/ no landslide). The Hosmer-Lemeshow statistics is a Chi-square Pearson statistics computed via the 2×g table of observed and expected frequencies. In such a table, 'g' shows the number of groups obtained from the estimated probability. In this test, the high p-value justifies the conformity of the model to the data on hand (Peng et al. 2002). The descriptive criteria to assess goodness of fit include R^2 indices introduced by Cox and Snell (1989) and Nagelkerke (1991). In linear regression, these indices have been well defined and their values indicate the ratio of the dependent variable changes that can be predicted and explained by the model. Since justification of changes observed in indices in logistic model is vague, and since they cannot be tested using an inferential method, it is recommended that these two indices be used as supplementary to fit test statistics (Peng et al. 2002). Further, recent studies have shown that RL^2 (The likelihood ratio R^2) is very useful in the logistic method (Menard, 2002). The Omnibus test is a likelihood-ratio Chi-square test that compares the current model with the zero model. In the zero model, the results of the analysis are reported without using any of the model's independent variables. This test shows the degree of success of the main model compared to the zero model (Pallant, 2007). The -2 Log likelihood statistics, behaving like Chi-square test, was also used to assess the model. The amount of this statistics for the logistic model, incorporating only an intercept, can be obtained by SPSS and by adding the Chi-square for the modeling of Omnibus Tests of Model Coefficients table plus the -2 log likelihood in the Model Summary table. If the amount obtained is big, the model shows low conformity, and if the amount is small, it shows that the model has good conformity with the data under study (Menard, 2002).

Besides, to validate the predicted probabilities, classification table in SPSS output was made use of in which the degree of the ability of the model in predicting the correct category (landslide/no landslide) has been shown for each cell.

There is another interesting model, namely Receiver Operating Characteristic (ROC), to assess fitness of the logistic model. In this method, a percentage of the observations (cells) with landslides – that have correctly been predicted by the model – are called Sensitivity (Probability for correctly identifying a positive or the true positives) based on Eq. 8.

$$\text{Sensitivity} = \frac{n_{tp}}{n_{tp} + n_{fn}} \tag{8}$$

n_{tp}: Number of true positive decisions,n_{fn}: Number of false negative decisions

Further, the specificity of the model has been shown based on the percentage of the correct classified observations (cells) with no landslides (Probability of correctly identifying a negative or true negatives) (Eq. 9).

$$\text{Specificity} = \frac{n_{TN}}{n_{TN} + n_{FP}} \qquad (9)$$

n_{TN}: Number of true negative decisions, n_{FP}: Number of false positive decisions.

The ROC is a very useful method to assess models that classify regions to landslide/no landslide zones (Swets, 1988; Zweig and Campbell, 1993; Yesilnacar and Topal, 2005; Mathew et al. 2009). Commonly, sensitivity takes the y-axis and (1– specificity) the x-axis. The Area under the ROC curve shows the increase in the probability of making a positive choice compared to a negative one. The ideal curve model has the maximum amount of the under curve layer. The total under curve area varies between 0 and 1. If the model used predicts the probability of landslide better than a random method, the under curve layer can be equal to 0.5. If the under curve layer in a model is 1, it would imply that the model has made the best and the most comprehensive predictions.

CAUSATIVE FACTORS

Geology

The type of lithology and geological structures are considered to be one of the most important factors in landslide occurrence (Lan et al. 2004; Ayalew and Yamagishi, 2005; Saldivar-Sali and Einstein, 2007). Due to variations in the geological formations of the region and various degrees of sensitivity of the rocks to landslide, they play an important role in spread of landslides in the region (Mahdifar, 1996; Talaei et al., 2004; Uromeihy and Mahdavifar, 2000). The rocks are late Precambrian to Recent (Davise et al. 1975) and could be divided into a number of distinct group namely calcareous, plutonic, volcanic and pyroclastic, metamorphic rocks as well as sedimentary rocks of Neogene and Quaternary age (Fig. 2). About 13.29% of the whole area is of Miocene clay to marly sedimentary rocks (Ng^{ml}) and 42.97% of altered zones of the area (a) had experienced landslides. This rate of landslide is remarkable compared to areas with other lithologies. Density of joint system, fractures, faults and crush zones can play an important role in slope instability (Haeri and Samee, 1997; Fatemi Aghda et al. 2003; Rutela and Lakhera 2000; Lan et al. 2004). The distance between the cells and the main faults varies from 0 to 23233 meters. The study has shown that 75.4% of the landslide zones occur within 0 to 8 kilometers from major faults (Fig. 3). Landslide

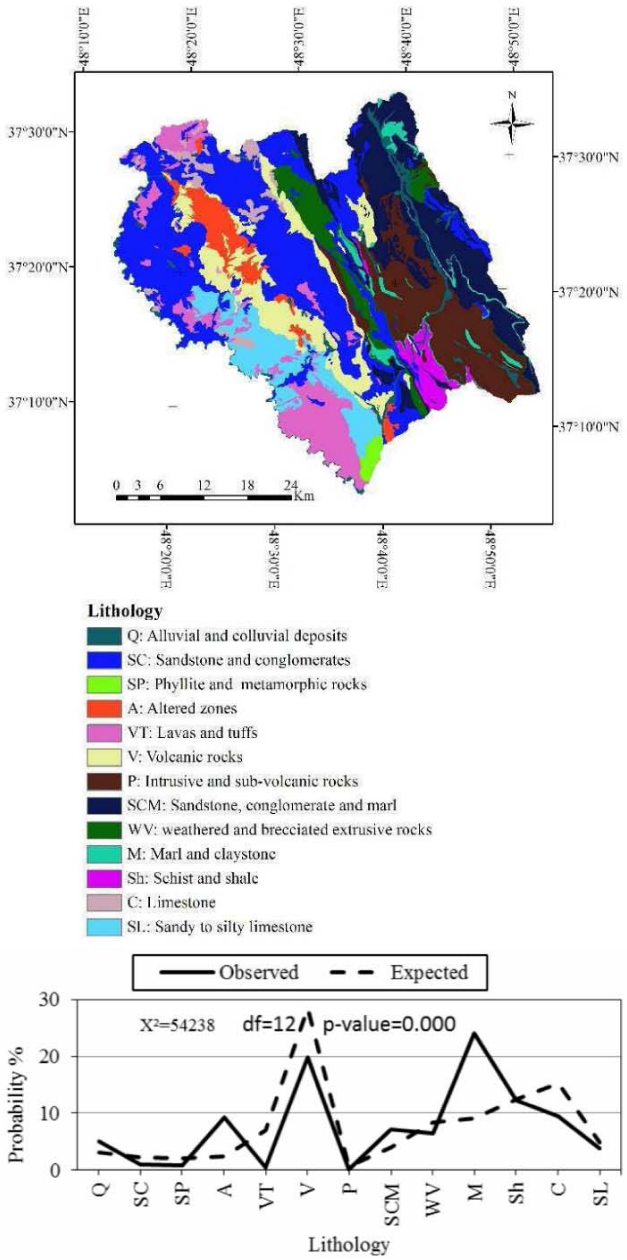

Fig.2. Lithological map of study area and landslide probability histogram and chi-square test result

occurrence is mostly due to the indirect impact of major and minor faults – resulting in the crushing of the surrounding rocks, increase in the permeability of the rocks, hydro-thermal activity and expansion of alteration zones. Further, faults can produce landslides through generation of earthquakes. In fact, 98.8% of the past landslides have occurred in areas with a Peak Ground Acceleration (PGA) of 0.57 to 0.58 (g).

Land Cover - Land Use Factor

The impact of plantation on slope stability occurs in a

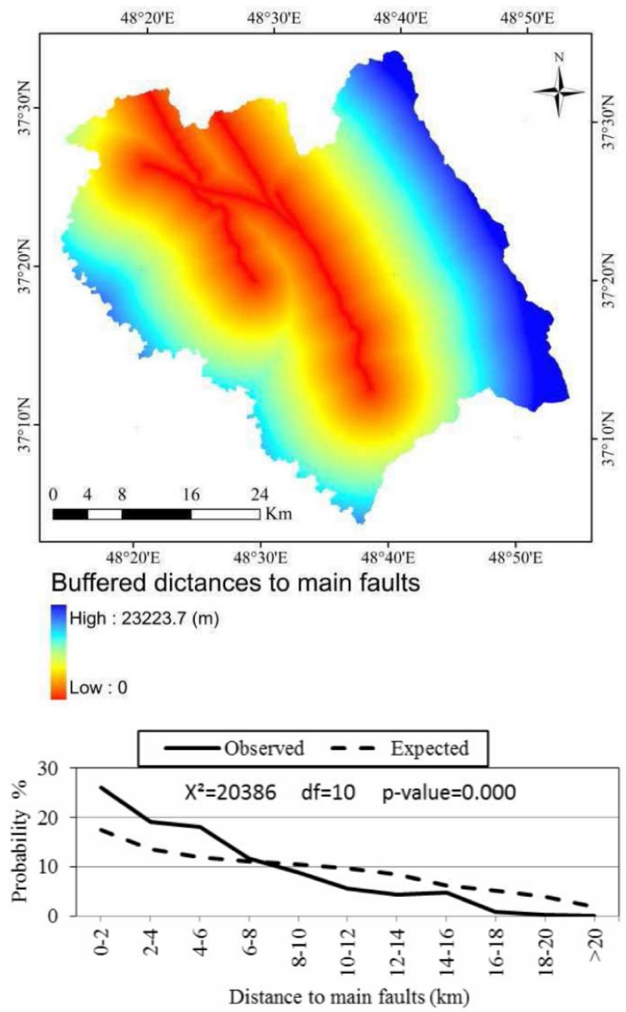

Fig.3. Buffer map of main faults and landslide probability histogram and chi-square test result

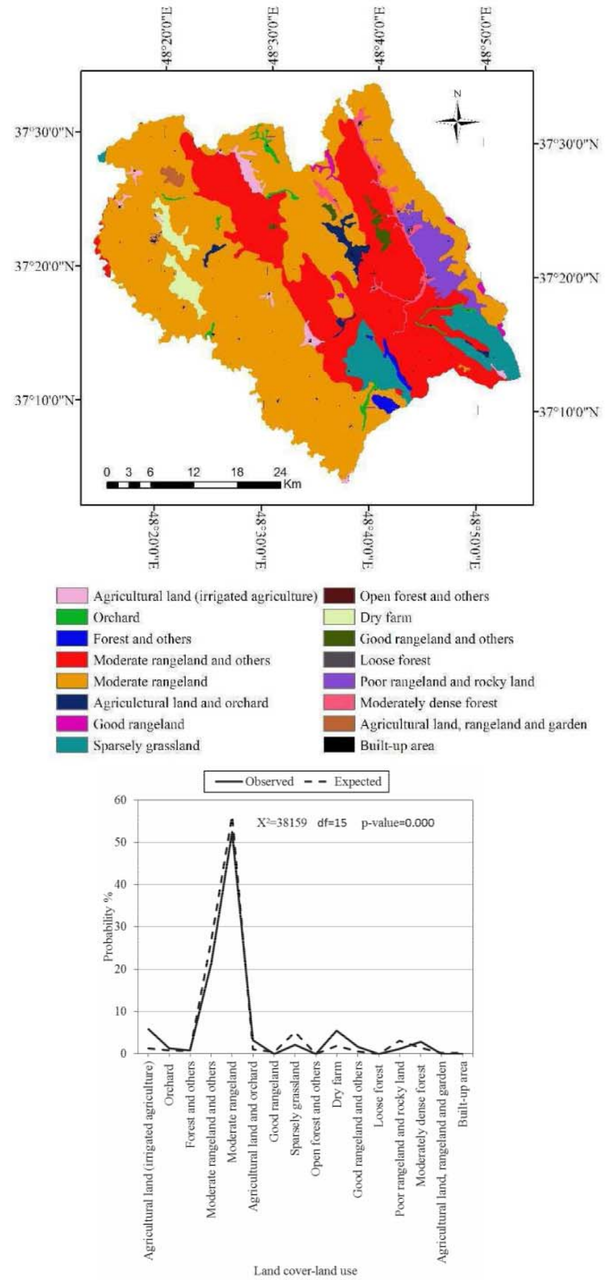

Fig.4. Land cover map of study area and landslide probability histogram and chi-square test result.

variety of ways including changes in cohesion, internal friction angle, weight of the soil and pore-water pressure (Gomez and Kavzoglu, 2005; Donati and Turrini, 2002).

The type of plantation and any changes therein can cause landslide (Mehrotra et al. 1996; Thomas, 2003; John et al. 2006). Land use in Hashtchin region is mainly water and rainfed agriculture, gardens, pasture and forest (Fig. 4). In the area under study, during the past 50 years, destruction of greenery, deforestation and shifting of land use to agriculture fields and gardens have increased the number of landslides in the region at 2 to 6 times. Comparison of the rate of landslides occurred (observed) and estimated (expected) shows that in agricultural farms and fields, landslide probability has increased 3 to 5 times. This has been up to 1.62 times in gardens, and 2.9 to 5.45 times in weak or changed pastures. Up to 300 meters of the roads and residential areas, no meaningful change is observed in the number of cells with landslides. Statistically speaking, a reduction in the distance of cells to residential places and roads (both with mostly low quality) does not show a meaningful impact of distance on past landslides. Nevertheless, due to the occurrence of a number of landslides in villages, roads, and parts of cities, it seems necessary to include the information on roads and residential places in analysis of slope of landslide susceptibility.

Geomorphology

Slope gradient and slope aspects, the amount and direction of runoff, plantation density, soil moisture and temperature are among the most important factors instigating landslides (Lan et al. 2004; Ayalew and Yamagishi, 2005; Gomez and Kavzoglu, 2005). Landslide probability in northeast-facing slopes is about 2.25 times higher than the expected rates for these slopes. Slopes with northeast and east orientations possess high landslide densities since they meet the Caspian sea cycles and accordingly has more humidity and rain. Landslide susceptibility is, thus, expected to increase with an increase in the degree of slope (Gomez and Kavzoglu, 2005). In the area under study, 85.68% of the grid-cells involved in landslides could be categorized into three slope levels namely level I, II, and III (Fig. 5). Lithology and the degree of slope are not independent of each other (Saldivar-Sali and Einstein, 2007). In many regions, an increase in the degree of slope cannot, on its own, control landslide occurrence (Duman et al. 2006). The slopes having claystone, mudstone and marly formations and altered zones of the region – with a slope degree of

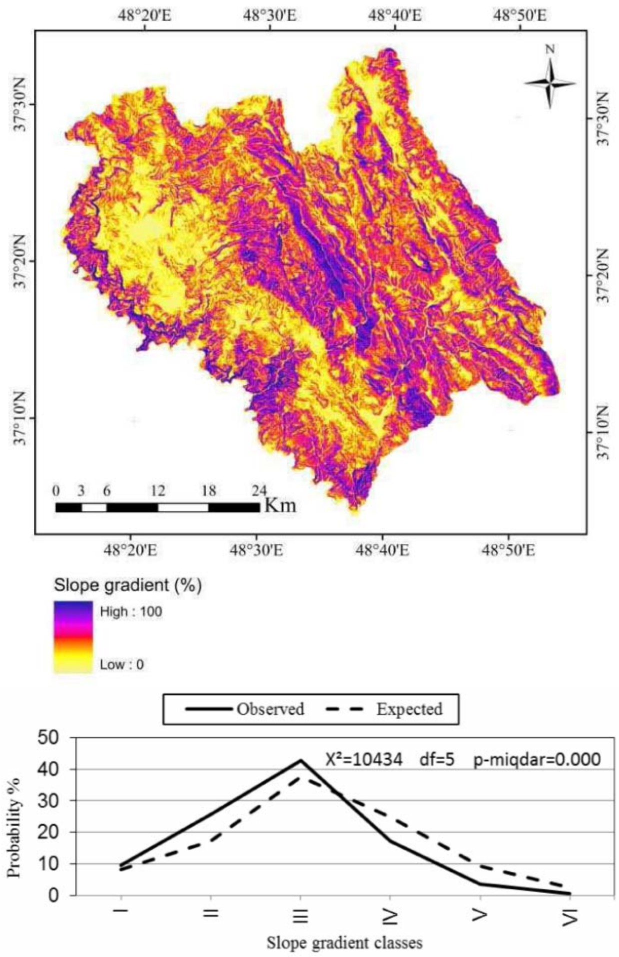

Fig.5. Slope gradient map of study area and landslide probability histogram and chi-square test result (I<10%; II=10-20%; III=20-30%; IV=30-40%; V=40-60%; VI>60%)

10% to 40% – reveal the important role of lithology in landslides of the region. In slopes with sensitive lithology and gentle slope, water permeability plays an important role in widespread landslides (Dai et al. 1999; Wong et al. 1998). The elevation of Hashtchin region varies between 850 meters in southeast, by the side of Qezel Owzan and 3324 meters in Agh Dagh. The probability ratio of the number of observed landslides to those predicted showed an increase, equal to 1.14 to 1.68 times, in areas with a height of 1000 to 2000 meters. This ratio was less than 1 in other areas. Such factors provide good conditions for permeability of rain into the underground. This condition can also initiate landslides (Kamp et al. 2008). To determine the impact of slope morphology on landslides, the slopes were divided into three groups namely flat, convex and concave. An analysis of the probability ratio of cells with landslide with predicted cells revealed that slope curvature in Hashtchin region does not have a meaningful impact on landslide occurrence.

Climatic and Hydrologic Factors

In some landslide prone areas, the mean annual precipitation is considered to be an important factor in landslide occurrence (Dai et al. 2002; Okamoto et al. 2004; Lan et al. 2004). The degree of the impact of rain on slope instability depends on the climatic conditions, geological structures, topography and permeability along slope formations (Haeri and Samiee, 1997). A comparison of the number of cells in regions with different rates of annual rain reveals that in the area under study about 65.5% of the cells with landslide have occurred in areas with mean annual precipitations of 300 to 360 mm. Since landslide probability does not increase with an increase in the amount of rainfall, landslide occurrence cannot be justified and explained directly through the amount of rainfall. In fact, factors like the amount of runoff and changes in the condition of groundwater level and flow can cause landslides since they change the mechanical characteristics of rocks and soil (Uromehyie and Mahdavifar, 2000). Therefore, the impact of rainfall on enhancing landslides susceptibility can be dependent on the geological conditions of slopes.

Groundwater aquifers are mainly found in calcareous sedimentary rocks in the eastern parts as well as in non-carbonate sedimentary strata comprising marl, mudstone, siltstone, sandstone, conglomerate, and volcanic rocks which comprise lavas and tuffs in the western part of the region. Being a mountainous area the groundwater table is highly variable in the study area. In general, the direction of groundwater flow is from the mountains to Qezel Owzan river in the western and southwestern boundary of the area. Because of high permeability, calcareous sedimentary rocks in the area have great potential to reserve water. Igneous and metamorphic rocks with joints and fractures developed due to folding and faulting, have been considered as semi-permeable units with underground water reserves. Other formations, such as the Neogene deposits of marl, mudstone, siltstone, sandstone and conglomerate are impermeable units and groundwater reserves in them are low. Therefore, with respect to the relative permeability, the lithological units in the study area can be divided into three groups: permeable, semi-permeable and impermeable (Pars Karst Water (P.K.W.), 2011) (Fig. 6). Statistical evaluation of the relationship between landslide occurrences and lithology shows that most landslides have occurred in impermeable areas (Table 1).

Table 1. Landslide densities of the relative permeability classes in the study area

Landslide density (%)	All grid (pixels) cells		Grid cells (pixels) with landslide		Relative permeability
	%	Frequency (no. of pixels)	%	Frequency (no. of pixels)	
13.64	59.2	419072	73.8	57164	Low or impermeable
6.8	20.7	146907	12.9	10008	Moderate (semi-permeable)
7.2	20.1	142477	13.3	10275	High (permeable)

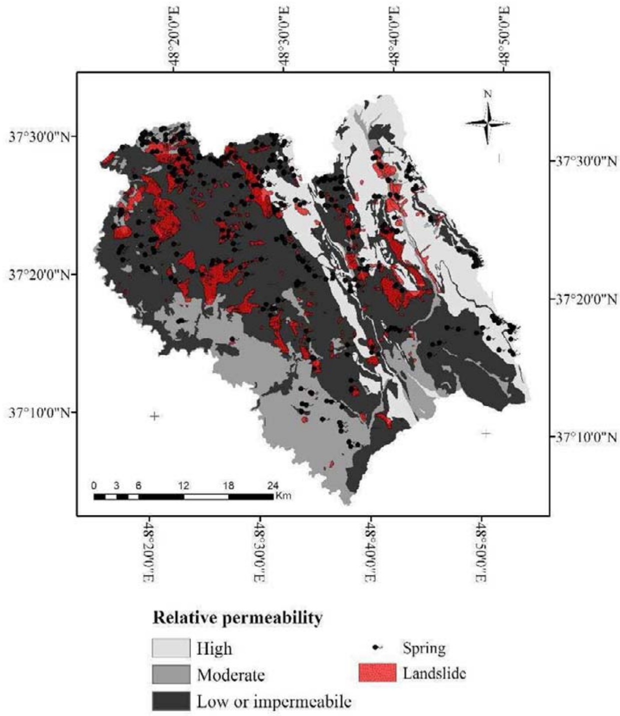

Fig.6. Relative permeability map of the lithologies.

The reason for this can be attributed to the springs because the majority of springs in the area discharge along the contact points of lithology with different permeability. During the field studies carried out for the preparation of landslide inventory maps, it was found that most springs that led to slope instabilities in the region are located at the boundary between the impermeable sedimentary units of Neogene age (mostly marls, mudstone, siltstone, sandstone and conglomerate) and semi-permeable igneous rocks with joints and fractures of Eocene age. Field check indicates that the failures generally occurred along the contact surfaces of sedimentary-volcanic rocks. Rapid changes in water levels during late winter or early spring rainfall caused new landslides or reactivation of the old landslides.

The main river in the area is Qezel Owzan which is in the western border of the area. Qezel Owzan river is the main branch of Sefid Rud basin, which has an area equal to 60,496 km^2 and ultimately discharges into the Caspian sea. Qezel Owzan is so long that it straddles through several natural regions or zones extending from Zagros to Alborz. Flow discharge of this river at Gylvan station (40 km after the area) is 111 m^3/s (Mahdavifar, 1996). Other major streams in the region include, Shah Rud, Zal, Saqezchi, Goolgoolab and Kandiraq which flow in west and southwest of the area and all of them pass into the Qezel Owzan channel. Although due to lack of stream-gauging stations discharges of these streams cannot be directly determined, but measuring Qezel Owzan river discharge before entering the area (Astor station, 20 km before entering the area) and after exiting it (Gylvan station, 40 km after the area) shows a 19.33 m^3/s increase in flow rate of which is partly because of aforementioned streams.

An increase in the amount of rainfall together with the increase in river discharge and acceleration in erosion along river banks that causes undercutting are all considered as important factors enhancing slope failure (Wu et al. 2004; Fourniadis et al. 2007). In Hashtchin area, stream bank erosion has been responsible for 81% of single landslides and 42% of landslide zones. In all, 53.3% of cells with landslide have been 0 to 900 meters off the stream bed.

The Input Data in Analysis of Logistic Regression

In order to estimate coefficients for logistic regression, it is necessary to have suitable input data. In fact, for each cell, quantities of independent and dependent (landslide/no landslide) variables must be available. In the present study, using a random selection, and for analytical purposes, 75% of cells (pixels) with landslide were used to estimate landslide susceptibility and the remaining 25% cells have been retained for accuracy assessment. From the produced database, 58054 cells with landslide and 58054 non-landslide were randomly selected for logistic regression analysis. In all, 19393 cells with landslide were kept for model validation. This combined spatial database consists of 116108 cells. In the first column of this tabular database, landslide occurrence (dependent variable) in the past was shown, in each cell, with 1 and no landslide occurrence was shown with 0. In other columns, independent variables data have been shown as string-nominal (categorical predictors) and numerical-scale (continuous predictors) data.

RESULTS

The first logistic model was designed based on Table 2 variables and 75% of the landslides of the region. This model was implemented with forward stepwise analysis using maximum likelihood method in SPSS Version 18. The regression started with 62 predictor variables units including continuous predictors and categorical predictors (Table 2). Since in this analysis, the independent variables, lithology, land cover and land use, slope aspect, altitude and curvature are all categorical, to differentiate categories dummy variables made by SPSS was used. For the categorical variables, the last category was selected as the reference with which all the other groups were compared. Hence, the number of dummy variables in each categorical variable is equal to the number of groups minus one (k, k-1 dummy variables categories). The simulated variables were used to compare the differences observed between the groups and the dependent variable (landslide/non-landslide).

There are several methods to select the variables to be included into the model. In this study, the forward stepwise method was adopted. In this method, score statistics were used to select variables as input into the model. This method started without independent variables, and finished having added the variables in ten steps. In each step, the predictor variable, the significance level of which is less than the specified value, default 0.05, was incorporated into the model. In the last step, the variables with significance values larger than 0.05 were excluded from the analysis. Those variables that could cause significant changes in the -2 log-likelihood were selected for analysis (Davis and Ohlmacher, 2002). The forward stepwise method of variable selection was stopped after completion of the 10^{th} step. Having deleted 37 predicted variable classes from the variables that had not caused any significant changes in -2 log-likelihood, 32 of variable classes were retained in the model. Since changes in -2 log-likelihood are more reliable than those in Wald statistics, the variables were selected using the -2 log-likelihood method (Table 3). The effect of each predictor variable has been summarized in parameters estimation table (Table 3). For each selected variable, the estimated amount for α_i coefficients has been shown. The null-hypothesis here is that the amount of the coefficient is equal to zero. To test the null-hypothesis, Wald statistics with the relevant degree of freedom was used. The selected variables with their relevant estimated coefficients had a significance level of less than 0.05 and were significantly different from zero, that is, the selected variables were effective in the model. Having estimated the constants and coefficients of the independent variables in the logistic regression analysis, in the next step, the probability ratios for all Hashtchin cells (50 m×50 m) were computed based on quantities of independent variables. This estimation was made based on the function defined as f(z) (Eq. 4). In Eq. 4, for each cell (pixel) in the study area, the continuous independent variable quantities are multiplied with the relevant coefficient. A cell with a categorical variable has been shown with the digit 1 and has been multiplied into its equivalent coefficient. Variables and independent classes not included in the model, have been shown as 0 and have been excluded from the analysis. For each cell, the result of the multiplication of the coefficients and quantities of the independent variable has been added together. The ultimate number was added to the constant b0. Then, landslide susceptibility was computed. The result has been shown in the form of a raster layer in which estimated probability quantities have been specified for each cell. This probability rate varies between 0 and 0.9954 (Fig. 7).The rate of cell or Hashtchin area susceptibility to landslide is determined based on the minimum and maximum probability levels obtained.

Landslide Susceptibility Map Grading

Landslide susceptibility map grading was undertaken by two methods: (1) standard deviations of landslide occurrence probability, and (2) dividing landslide susceptibility rates to four equal groups (Fig. 8 and 9). Later, the results obtained in each method were assessed. In the first method, 38.6% of the area under study was placed in the group with very low susceptibility. Low and moderate susceptibility classes comprised of 28.9% and 13.2% of the area, respectively. In all, 9.4% of the region was placed in the class with very high landslide susceptibility (Table 4).

Accuracy Assessment

Landslide susceptibility zonation map of Hashtchin region was prepared using logistic regression model in SPSS, Version 18. The first output of this method is a zero block, also headed beginning block, in which statistical analysis is carried out without using any independent variables. The results obtained were used for the purpose of comparison with the model where independent variables are included.

Table 2. List of independent variables used in logistic regression analysis

Sl. no.	Variable	Nature	Class code	Description
1	Lithology	Categorical	Class I	Q: Alluvial and colluvial deposits
			Class II	SC: Sandstone and conglomerates
			Class III	SP: Phyllite and metamorphic rocks
			Class IV	A: Altered zones
			Class V	VT: Lavas and tuffs
			Class VI	V: Volcanic rocks
			Class VII	P: Intrusive and sub-volcanic rocks
			Class VIII	SCM: Sandstone, conglomerate and marl
			Class IX	WV: weathered and brecciated extrusive rocks
			Class X	M: Marl and claystone
			Class XI	Sh: Schist and shale
			Class XII	C: Limestone
			Class XIII	SL: Sandy to silty limestone
2	Slope aspect	Categorical	Class I	N facing (45° about N)
			Class II	NE facing (45° about N 45° E)
			Class III	E facing (45° about E)
			Class IV	SE facing (45° about S 45° E)
			Class V	S facing (45° about S)
			Class VI	SW facing (45° about S 45° W)
			Class VII	W facing (45° about W)
			Class VIII	NW facing (45° about N 45° W)
			Class IX	Flat
3	Altitude (Elevation) (m)	Categorical	Class I	400=TEL=600
			Class II	600=TEL=800
			Class III	800=TEL=1000
			Class IV	1000<TEL=1200
			Class V	1200<TEL=1400
			Class VI	1400<TEL=1600
			Class VII	1600<TEL=1800
			Class VIII	1800<TEL=2000
			Class IX	2000<TEL=2200
			Class X	2200<TEL=2400
			Class XI	2400<TEL=2600
			Class XII	2600<TEL=2800
			Class XIII	2800<TEL=3000
			Class XIV	3000<TEL=3200
4	Topographic curvature	Categorical	Class I	Concave (-)
			Class II	Flat (0)
			Class III	Convex (+)
5	Land cover	Categorical	Class I	Agricultural land
			Class II	Orchard
			Class III	Forest and others
			Class IV	Moderate rangeland and others
			Class V	Moderate rangeland
			Class VI	Agricultural land and orchard
			Class VII	Good rangeland
			Class VIII	Sparsely grassland
			Class IX	Open forest and others
			Class X	Dry farm
			Class XI	Good rangeland and others
			Class XII	Loose forest
			Class XIII	Poor rangeland and rocky land
			Class XIV	Moderately dense forest
			Class XV	Agricultural land, rangeland and garden
			Class XVI	Build-up area
6	Distance to main faults (m)	Continuous		
7	Distance to drainage (m)	Continuous		
8	Slope gradient (°)	Continuous		
9	Distance to road (m)	Continuous		
10	Distance to settlement (m)	Continuous		
11	Peak ground acceleration (g)	Continuous		
12	Mean annual precipitations (mm/year)	Continuous		

Table 3. The predictors retained in the final logistic regression model and their estimated coefficients

Variables	b_i	S.E.	Wald	df	Sig.	$Exp(b_i)$
Distance to main faults (m)	.000	0.000	2834.863	1	0.000	1.000
Peak ground acceleration (PGA) (g)	262.778	3.939	4450.923	1	0.000	92.23×10^{14}
Mean annual precipitations (mm/year)	.008	0.000	1026.902	1	0.000	1.008
Distance to road (m)	.000	0.000	748.772	1	0.000	1.000
Distance to settlement (m)	.000	0.000	606.942	1	0.000	1.000
Slope gradient (°)	-0.011	0.000	632.383	1	0.000	.989
Lithology			7941.667	12	0.000	
Class I	2.962	.082	1289.915	1	0.000	19.336
Class II	1.304	0.071	333.170	1	0.000	3.683
Class III	10.942	0.200	2978.516	1	0.000	56510.290
Class IV	1.563	0.077	415.802	1	0.000	4.773
Class V	2.451	0.076	1050.120	1	0.000	11.602
Class VI	2.309	0.072	1020.544	1	0.000	10.064
Class VII	2.322	0.075	949.713	1	0.000	10.192
Class VIII	2.104	0.076	762.408	1	0.000	8.196
Class IX	1.957	0.080	599.621	1	0.000	7.078
Class X	1.554	0.091	294.764	1	0.000	4.730
Class XI	1.380	0.100	191.309	1	0.000	3.975
Class XII	3.347	0.083	1623.310	1	0.000	28.424
Slope aspect			712.432	8	0.000	
Class III	.133	0.031	18.002	1	0.000	1.142
Class V	-.481	0.032	227.339	1	0.000	0.618
Class VI	-.380	0.030	159.064	1	0.000	0.684
Class VII	-.243	0.029	70.308	1	0.000	0.784
Class VIII	-.307	0.029	109.681	1	0.000	0.736
Land cover-land use			4307.086	13	0.000	
Class I	0.954	0.080	143.129	1	0.000	2.597
Class II	0.956	0.068	195.291	1	0.000	2.602
Class III	0.281	0.093	9.031	1	0.003	1.324
Class IV	-1.810	0.073	621.196	1	0.000	0.164
Class V	-0.242	0.071	11.657	1	0.001	0.785
Class VI	0.849	0.133	40.623	1	0.000	2.338
Class VIII	-0.225	0.087	6.724	1	0.010	0.798
Class X	-1.084	0.050	471.197	1	0.000	0.338
Class XI	-0.409	0.049	68.655	1	0.000	0.664
Constant	-199.95	10329.0	0.000	1	0.985	0.000

SE: Standard error of estimate, Wald: Wald chi-square values, df: degree of freedom, Sig: significance, $Exp(b_i)$: exponentiated coefficient

The degree of preliminary success in this analysis (block 0) has been presented in the classification table (Table 5). This table shows that 100% of the cells with landslide have been predicted accurately, further, none of the cells with no landslide were predicted accurately. At this stage, the accuracy of the model in landslide prediction was 50%.

Table 6 shows the degree of the success of the model with all of the predictors entered into the model, in predicting landslide/no landslide situations in each cell. The rate of improvement in the main model's prediction ability could be determined through comparison of the results obtained in this table with those earlier presented in Table 6 (the results obtained for block 0, with none of the predictors entered into model). In the 10^{th} step the inclusion of collective independent variables into the model, 76.5% of the cells were accurately predicted. This rate is very satisfactory given

Table 4. Descriptive statistics of landslide susceptibility maps grading using the two methods of equal distances and standard deviations

	Standard deviation			Equal interval		
Class name	Frequency Percent	Percent	Cumulative	Frequency	Percent	Cumulative Percent
Very low	273329	38.6	38.6	290911	41.1	41.1
Low	204511	28.9	67.4	185829	26.2	67.3
Medium	164055	23.2	90.6	153177	21.6	88.9
High	66561	9.4	100.0	78539	11.1	100.0
Total	708456	100.0		708456	100.0	

Table 5. The classification table (Step 0)

Observed	Predicted		
	No landslide	Landslide	Percentage correct
No landslide	0	58054	0.0
Landslide	0	58054	100.0
Overall percentage	50.0		

Constant is included in the model and the cut-off value is 0.500

Table 6. The classification table (Step 10, block 1)

Observed	Predicted		
	No landslide	Landslide	Percentage correct
No landslide	42435	15619	73.1
Landslide	11679	46375	79.9
Overall percentage		76.5	

Constant is included in the model and the cut-off value is 0.500

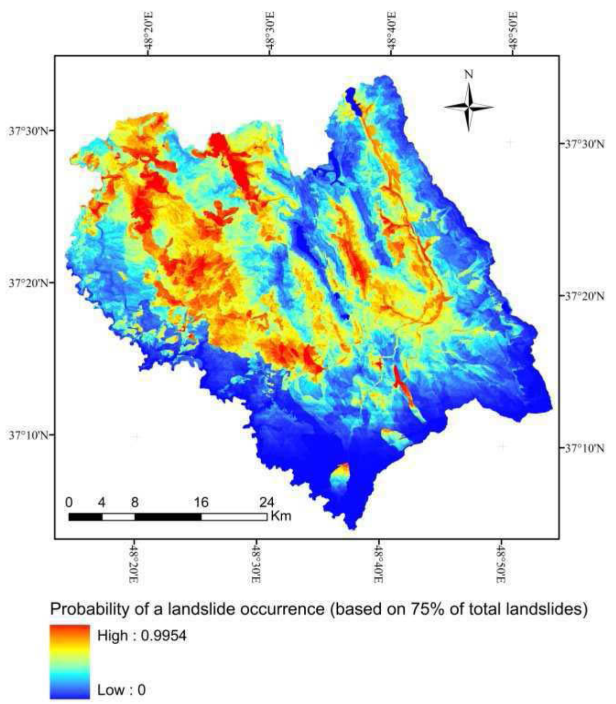

Fig. 7. The susceptibility map of landslide occurrence using the logistic regression approach

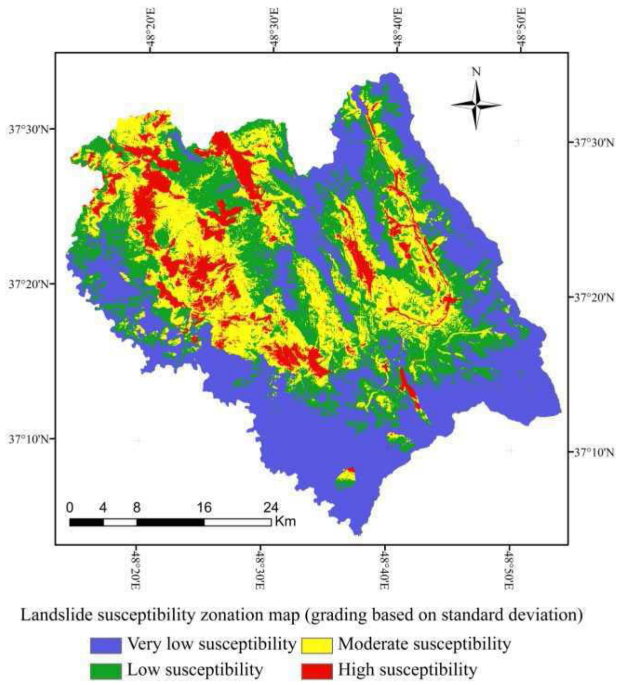

Fig.8. Map of relative landslide susceptibility using the standard deviation method

the 50% rate obtained in the zero model. In this model, 46357 cells out of the whole 58054 cells with landslides, and 42435 cells out of the whole 58054 cells with no landslides have been accurately classified. Based on the findings it could be concluded that the forward stepwise logistic model produces an appropriate model, at its 10th step, for prediction of landslide/no landslide in each cell.

The Omnibus test with a significance value of less than 0.05% (which really means $p<0.0005$), shows the good performance of the model compared to the zero model (Table 7). Therefore, the model with our set of independent variables used as predictors (the effective factors in the landslide susceptibility of Hashtchin region) performs much better than the presupposed model performed in the beginning block (block 0).

The statistical values computed for -2 log likelihood, Cox and Snell (1989) and Nagelkerke (1991), in the 10th step, show that 33.1% and 44.2% of the amount of variation in the dependent variables can well be justified and explained by this model. The categorical variables reduce such statistics value (Table 8). As the model moves on from one step to the next, the statistics computed for -2 log-likelihood decreases. This is a good indicator of its suitability and fitness of the model.

The percentage of the cells with landslide that have been accurately predicted is labeled as the indicators of sensitivity of model (true positives). The model proposed is able to classify 79.9% of the cells with landslide accurately. The specificity of the model is shown based on the percentage of the whole cells with no landslide that have been classified accurately. The proposed specificity of model was 73.1%. In other words, the present model could classify accurately 73.1% of the cells with no landslide. In Fig. 10, the ROC curve of the implemented model is shown. The area under the ROC curve (AUC) is 0.841, equivalent to an accuracy of 84.1%, which is a very good for landslide prediction.

Table 7. Results of the model's Omnibus test in variables input steps

		Chi-square	df	Sig.
Step 10	Step	612.803	1	.000
	Block	46705.032	53	.000
	Model	46705.032	53	.000

Table 8. Results of testing the model

Step	-2 Log likelihood	Cox and Snell R Square	Nagelkerke R Square
1	143107.348[a]	.143	.190
2	137019.249	.186	.248
3	129556.275	.237	.316
4	120373.767	.295	.393
5	117664.844	.311	.415
6	116550.954	.318	.424
7	115897.579	.322	.429
8	115294.386	.325	.434
9	114867.637	.328	.437
10	114254.833	.331	.442

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

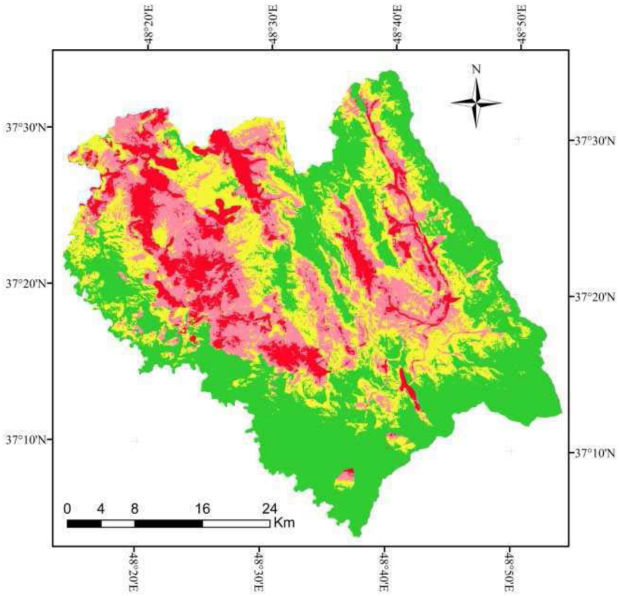

Fig.9. Map of relative landslide susceptibility using the equal distances method.

The asymptotic significance is less than 0.05, which means that using the assay is better than guessing (Table 9).

Assessment of Different Sensitivity Degrees of Susceptibility Maps

To assess the degree of fit of the classified susceptibility map based on the methods mentioned above, the Eq. 7 was used (FernÜndez et al. 2003). The smaller the degree of fit in classes with very low and low landslide susceptibility (relative error) and the higher in classes with moderate to high landslide susceptibility (relative success rate) shows the high quality of the susceptibility map. Further, zonation maps also could be assessed through determining the failure percentage of zonations in each susceptibility class. In this latter method, estimating the absolute error and the degree of success (in order of the failure percentages in very low, low, moderate and very high susceptibility classes) is also possible. Based on the findings and the data obtained in Tables 10 and 11, the following conclusions could be drawn:

- In both susceptibility grading methods – equal intervals and standard deviation – the rate of error always falls below 10%. Of course, the rate of error in the standard deviation method has been relatively lower than that obtained in the method based on the equal intervals.
- Broadly speaking, zonation maps produced by either method have proved highly qualitative and could be used as the basis for further studies.

Table 9. The area under the curve test results

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.841	.001	.000	.840	.842

a. Under the nonparametric assumption; b. Null hypothesis: true area = 0.5

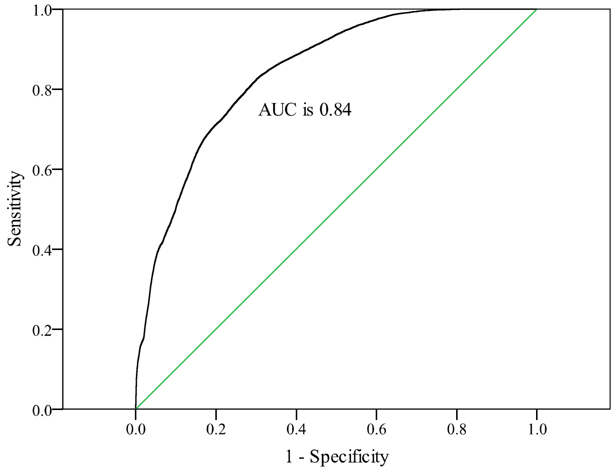

Fig.10. Receiver Operating Characteristic curve (ROC) for logistic regression model based on the estimation model with 75% of all cells with landslide

DISCUSSION

The natural features of a region, e.g. geology, tectonic, lithology, climate and morphology condition, play a role for landslide occurrence. The reasons for occurrence of landslides are many which are complicated, and at times

Table 10. The degree of fit (%) of the susceptibility classes

Methods		Susceptibility class			
		Very low susceptibility	Low susceptibility	Medium susceptibility	High susceptibility
Std. Deviation classification	D.F.%	1.5	8.8	28	61.6
Equal interval classification		1.72	9.56	27.65	61.06

Table 11. The percentage of the rupture zones that lay within each susceptibility class

Methods	Percentages of the rupture zones that lay within each susceptibility class			
	Very low susceptibility	Low susceptibility	Medium susceptibility	High susceptibility
Std. deviation classification	3.69	16.59	42.09	37.60
Equal interval classification	4.42	15.71	37.45	42.40

unknown. Although the main factors in causing landslides can be identified through field studies or aerial and satellite images, some remain covert and unknown. To determine unknown factors, a number of geomorphometric parameters have been considered. Some of these geomorphometric factors are subjective and their conversion into quantitative variables is complicated. The logistic regression method is able to ease or solve this problem. This method can delete all the irrelevant factors and determine the degree of importance of factors contributing to landslides. A landslide inventory map of the study area, as the first step, derived from topographic map (at the scale of 1:25000), aerial photograph interpretation and field checks shows 137 landslide locations and 38 zones with multiple landslides (Fig. 1c). This data together with the data related to effective factors were stored in the GIS database. The results obtained from the contingency tables revealed that the highest number and area of landslides occurred on Miocene clay to marl sedimentary rocks (upper red formations), and on altered volcanic formations (related to Eocene and Oligocene) due to the presence of clay minerals. Besides the lithological factor, the largest number of landslides occurred in areas which are used for garden (9715.53 hectares), 6 to 18 kilometers off the main faults (8047.5 hectares), in the slope segments oriented to the northeast (2000 hectares) and areas with mean annual precipitations of 260 to 460 mm (13514.25). Gentle slopes also exhibit landslides with a slope degree of 5% to 20% (9157.5%), where lithology is more important causative factor than of the slope steepness. Using this database, the landslide susceptibility analysis was carried out using the logistic regression method. In modeling the landslide susceptibility of region, a number of variables were entered into the model. These variables were selected based on the features and results of analysis of factors contributing to landslides in Hashtchin area. Lithology, land use, slope morphology (profile curvature), altitude (elevation), and slope aspect were selected as categorical variables, whereas variables of distance to main faults, slope gradient, the mean annual precipitations, distance to rivers and roads, Peak Ground Acceleration (PGA) due to earthquake and distance from the buildings were entered as continuous variables. Then, the forward step-wise method of the logistic regression model was run and within 10 steps, the variables, altitude, land use, and three slope aspects were excluded from the model. Logistic regression analysis has provided estimates of the constant and the coefficients of the independent variables. Based on the estimated coefficients (b_i) and (b_0), the relevant equation was defined with the help of Eq. 4. Based on the coefficients and equations obtained, for each cell in the study area, landslide probability was computed in each cell between 0 and 1.

Although logistic regression has been widely used in many landslide zones (Lee and Min, 2001; Lee 2004; Ayalew and Yamagishi, 2005; Can et al. 2005; Wang and Sassa, 2005; Yesilnacar and Topal, 2005; Duman et al. 2006; Mathew et al. 2009), Hashtchin landslide susceptibility zonation mapping using logistic regression has been used first in the present study. The main point of logistic regression model lies in the nature and type of the data input into the model. The input variables in logistic regression model are selected based on the data available and features of the region under study (Mathew et al. 2009; Wang and Sassa, 2005; Yeslinacar and Topal, 2005; Lee, 2004). Of course, the results obtained from a model cannot be trusted unless after its comprehensive assessment and evaluation. To assess the logistic regression model, most researchers have made use of statistical methods (Ayalew and Yamagishi, 2005; Yesilnacar and Topal, 2005; Mathew et al. 2009).

The results obtained from the statistical analysis, in this research, showed that the model had a good performance and its predictions were accurate in 76.5% of the cases. To show the quality of probability and prediction methods, the ROC curve has proved helpful (Swets, 1988). Based on the area under the ROC curve (AUC), the prediction ability of the model is 84.1%, which is good for prediction of the landslide/no landslide susceptibility in Hashtchin region. Finally, the degree of success of the model in landslide susceptibility maps was found to be 79%.

CONCLUSIONS

The results indicate that the logistic regression model is convenient and applicable for scale adopted in this research and can be applied successfully to landslide susceptibility zonation mapping in the study area. Such susceptibility zonation maps can play an important role in mitigating landslide hazards and helping decision makers to take appropriate measures and plan intended developmental activities in slopes with high susceptibility in the region. The map may also be used as a basis for the landslide risk assessment studies to be applied in the study area in the future. The model introduced can also be applied to mountainous areas with similar features including Alborz, Zagros and Caucuses Mountains.

Acknowledgment: My sincere thanks to the head and members of Arya Research Center (aryarc.com) for their support for publication of this article.

References

ALEOTTI, P. and CHOWDHURY R. (1999) Landslide hazard assessment: summary review and new perspectives. Bull. Engg. Geol. Environ., v.58, pp.21-44.

ANBALAGAN, D. (1992) Landslide hazard evaluation and zonation mapping in mountainous terrain. Engg. Geol., v.32, pp.269–277.

ANSARI, F. and BLURCHI, M.C. (1996) Landslides of Ardabile Province, Iran (in Persian). Geol. Surv Iran, Iran. 46 pp.

ATKINSON, P.M. and MASSARI, R. (1998) Generalized linear modeling of landslide susceptibility in the Central Apennines, Italy. Computers & Geosciences, v.24(4), pp.373-385.

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, v.65, pp.15-31.

AYALEW, L., YAMAGISHI, H., MARUI, H. and KANNO, T. (2005) Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. Engg. Geol., v.81, pp.432-445.

BAKKER, W.H., GRABMAIER, K.A., HUNMEMAN, G.C., VAN DER MEER, F.D., PRAKASH, A., TEMPFLI, K., GIESKE, A.S.M., HECKER, C.A., JANSEEN, L.L.F., PARODI, G.N., REEVES, C.V., WEIR, M.J.C., GORTE, B.G.H., HORN, J.A., KERLE, N., POHL, C., VAN RUITENBEEK, F.J. and WOLDAI, T. (2004) Principles of remote sensing, an introductory textbook. The international institute for geo-informational science and earth observation (ITC), Enschede, the Netherlands. 540 p.

BRABB, E.E., PAMPEYAN, E.H. and BONILLA, M.G. (1972) Landslide susceptibility in San Mateo County, California. Misc. Field Studies map MF360 (scale 1:62500). U.S. Geol. Surv., Reston, Va.

CAN, T., NEFESLIOGLU, H.A., GOKCEOGLU, C., SONMEZ, H.N. and DUMAN, T.Y. (2005) Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. Geomorphology v.72, pp.250-271.

CARRARA, A., CARDINALI, M., DETTI, R., GUZZETI, F., PASQUI, V. and REICHENBACH, P. (1990) Geographical information system and multivariate models in landslide hazard evaluation. *In:* ALPS 90: Alpine landslide practical seminar, sixth international conference and field workshop on landslides, Aug. 31-Sept.12, Milan, Italy, Universitá degli Studi de Milano, pp. 17-28.

CARRARA, A. (1983) Multivariate models for landslide hazard evaluation. Mathematical Geol., v.3, pp.403-426.

Carrara, A., Cardinali M., Detti R., Guzzetti F., Pasqui V., Reichenbach P., 1991. GIS techniques and statistical models in evaluating landslide hazard. Earth Surface Processes and Landforms, v.16, pp.427-445.

CHANG, K.T., CHIANG, S.H. and HSU, M.L. (2007) Modeling typhoon- and earthquake-induced landslides in a mountainous watershed using logistic regression. Geomorphology, v.89 (3-4), pp.335-347.

COX, D.R. and SNELL, E.J. (1989) The analysis of binary data. (2nd Ed.) Chapman and Hall, London, 236p.

DAI, F.C. and LEE, C.F. (2002) Landslide characteristics and slope instability modeling using GIS Lantau Island, Hong Kong. Geomorphology, v.42, pp.213-228.

DAI, F.C., LEE, C.F., LI, J. and XU, Z.W. (2001) Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. Environ. Geol., v.40(3), pp.381-391.

DAI, F.C., LEE, C.F. and WANG, S.J. (1999) Analysis of rainstorm-induced slide-debris flows on natural terrain of Lantau Island, Hong Kong. Engg. Geol., v.51, pp.279-290.

DAI, F.C., LEE, C.F. and NGAI, Y.Y. (2002) Landslide risk assessment and management: an overview. Engg. Geol., v.64 (1), pp.65-87.

DAVIES, R.G., CLARK, G.C., HAMZEPOUR, B. and JONES, C.R. (1975) Explanatory Text of the Bandar-e-Pahlavi quadrangle map, 1:250000. Geol. Surv. Iran, No. D3. 203p.

DAVIES, R.G., JONES, C.R., HAMZEPOUR, B. and CLARK, G.C. (1972) The Geology of the Masuleh Sheet (northwest IRAN). Geol. Surv. Iran, Rep. No. 24, 110p.

DAVIS, J.C., CHUNG, C.J. and OHLMACHER, G.C. (2006) Two models for evaluating landslide hazards, Computers & Geosciences, v.32, pp.1120-1127.

DAVIS, J.C. and OHLMACHER G.C. (2002) Landslide hazard prediction using generalized logistic regression. *In:* Proc. 8th Annual Conf. Internat. Assoc. Mathematical Geol., Berlin, Germany, pp. 501-506.

DONATI, L. and TURRINI, M.C. (2002) An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology: application to an area of the Apennines (Valnerina; Perugia, Italy). Engg. Geol., v.63, pp.277-289.

DONG J.J., TUNG, Y.H., CHEN, C.C., LIAO, J.J. and PAN, Y.W. (2011) Logistic regression model for predicting the failure probability of a landslide dam. Engg. Geol., v.117, pp.52-61.

DUMAN, T.Y., CAN T., GOKCEOGLU, C., NEFESLIOGLU, H. A. and Sonmez, H. (2006) Application of logistic regression for landslide susceptibility zoning of Cekmece Area, Istanbul, Turkey. Environ. Geol., 51, pp.241-256.

DWI WAHONO, B. F. (2010) Applications of statistical and heuristic methods for landslide susceptibility assessments: A case study in Wadas Lintang Sub District, Wonosobo Regency, Central Java Province, Inonesia. Supervisor: Sri Hadmoko, D. and van Westen C. J., Gadjah Mada University, International Institute for Geo-Information and Earth Observation. Thesis for the degree of Master of Science. pp. 106.

FARIDI, M. and ANVARI, A. (1996) Geological map of the Hashtchin area, 1:100000. Iran: Geol. Surv. Iran, No. 5664.

FATEMI AGHDA, M., GHAYOMIAN J., FARAHANI ASHGHALI A. (2003) Evaluation of statistical methods in landslide hazard analaysis. Geol. Surv. Iran, Geosciences, No.47-48, pp. 28-47.

FERNÁNDEZ, T., IRIGARAY, C., EL HAMDOUNI, R. and CHACÓN, J. (2003) Methodology for Landslide Susceptibility Mapping by Means of a GIS. Application to the Contraviesa Area (Granada, Spain). Natural Hazards, v.30, pp.297-308.

FOURNIADIS, I.G., LIU, J.G. and MASON, P.J. (2007) Landslide hazard assessment in the Three Gorges area, China, using ASTER imagery: Wushan–Badong. Geomorphology, v.84, pp.126–144.

Gómez, H. and Kavzoglu, T. (2005) Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. Engg. Geol., v.78, pp.11-27.

Greco, R., Sorriso-Valvo, M. and Catalano, E. (2007) Logistic regression analysis in the evaluation of mass-movement susceptibility: the Aspromonte case study, Calabria, Italy. Engg. Geol., v.89, pp.47–66.

Grozavu, A., Mărgărint, M.C. and Patriche, C.V. (2012) Landslide susceptibility assessment in the brãieºti-sineºti sector of Iaºi Cuesta. Carpathian Jour. Earth Environ. Sci., v.5(2), pp.61-70.

Guzzetti, F., Carrara, A., Cardinali, M. and Reichenbach, P. (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology, v.31, pp.181-216.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M. and Galli, M. (2006) Estimating the quality of landslide susceptibility models. Geomorphology, v.1-2, pp.166-184.

Haeri, S.M. and Samiee, A.H. (1997) A new method for landslide hazard microzonation based on micro-zonation studies in Mazandaran Province. Geol. Surv. Iran, Geosciences, v.23-24, pp.2-15.

Hashemi Tabatabaei, S. (1998) Landslide hazard zonation in southwest of Ardabil Province Iran (in Persian). Ministry of Roads and Urban Development, Tehran, Iran. Vol. 2. 72p.

Hemmati, R., Dolatimehr, A., Nasirifar A., Shahbazi, M, Hezhabrpour, Gh. and Aghaei Kh. (2007) Ardabil Province climate (in Persian). Applied Meteorology research center of Ardabil, Islamic Republication of Iran Meteorological Organization, Ministry of Roads and Urban Development, Iran, 143p.

Ilinca, V. and Gheuca, I. (2011) The red lake landslide (Ucigaºu Mountain, Romania). Carpathian Jour. Earth Environ. Sci., v.6(1), pp.263-272.

Jadda, M., Shafri, H.Z.M., Mansor, S.B., Sharifikia, M., and Pirasteh, S., 2009. Landslide susceptibility evaluation and factor effect analysis using probabilistic-frequency ratio model. European Jour. Scientific Res., v.33 (4), pp.654-668.

John, R.D., Anne-Gaelle, A., James, D.S. and Lavs, B. (2006) Validation of a region-wide model of landslide susceptibility in the Manawatu-Wanganui region of New Zealand. Geomorphology, v.1-4, pp.70-79.

Kamp, U., Growley, B.J., Khattak, G.A. and Owen, L.A., (2008) GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. Geomorphology, v.101, pp.631–642.

Kleinbaum, D.G. and Klein, M. (2010) Logistic regression: a self learning text. Third edition. Springer, New York. 701p.

Knapen, A., Kitutu, M.G., Poesen, J., Breugelmans, W., Deckers J. and Muwanga A. (2006) Landslides in a densely populated county at the foot slopes of Mount Elgon (Uganda): Characteristics and causal factors. Geomorphology, v.73, pp.149–165.

Komakpanah, A. and Hafezimoghadasi, N. (1994) Landslide hazard zonation methods (in Persian). Workshop proceeding on landslide damage mitigation strategies in Iran, International Institute of Earthquake Engineering and Seismology, pp. 390-414.

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. Engg. Geol., v.76 (1-2), pp.109-128.

Lee, S. (2004) Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS. Environ. Magmt., v.34 (2), pp.223–232.

Lee, S. (2005) Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. Internat. Jour. Remote Sensing, v.7, pp.1477-1491.

Lee, S. and Min, K. (2001) Statistical analysis of landslide susceptibility at Yongin, Korea. Environ. Geol. v.40, pp.1095-1113.

Mahdavifar, M.R. (1996) Landslide hazard zonation in Khorshrostam region (southwest of Khalkhal County) (in Persian). MS Diss., Tarbiat Modarres University, Tehran, Iran. 154p.

Mathew, J., Jha, V.K., Rawat, G.S. (2009) Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. Landslides, v.6, pp.17–26.

Mehrotra, G., Kanungo, D. and Mahadeviah, K. (1996) Landslide hazard assessment- A need for environmental management. Proc. 7th Internat. Symp. Landslides, Trondheim, Norway, pp.315-320.

Menard, S. (2002) Applied Logistic Regression Analysis, 2nd (Ed.). Sage University Paper Series on Quantitative Applications in Social Sciences, v.106. Thousand Oaks, California. 111p.

Meusburger, K. and Alewell, C. (2009) On the influence of temporal change on the validity of landslide susceptibility maps. Nat. Hazards Earth Syst. Sci. v.9, pp.1495-1507.

Nagelkerke, N.J.D. (1991) A note on a general definition of the coefficient of determination. Biometrika, v.78, pp.691–692.

Naranjo, J.L., Van Westen, C.J. and Soeters, R. (1994) Evaluating the use of training areas in bivariate statistical landslide hazard analysis: a case study in Colombia. ITC-Jour., v.3, pp.292-300.

Niemann, K.O. and Howes D.E. (1991) Applicability of digital terrain models for slope stability assessment. ITC-Jour., v.3, pp.127-137.

Nikandish, N., Mir Sanei R., 1996. Introduction to Ardabile Province landslides, Iran (in Persian). Ministry of Jihad-e-Agriculture, Tehran, Iran. 63 pp.

Ohlmacher, C.G., Davis C.J., 2003. Using multiple regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Engg. Geol., v.69, pp.331–343.

Okamoto, T., Larsen, J.O., Matsuura, S., Asano, S., Takeuchi ,Y. and Grande, L., 2004. Displacement properties of landslide masses at the initiation of failure in quick clay deposits and the effects of meteorological and hydrological factors. Engg. Geol., v.72, pp.233–251.

Pallant, J. (2007) A step-by-step guide to data analaysis using

SPSS version 15, SPSS, survival manual. McGraw-Hill, Australia. 335p.

Pars Karst Water (P.K.W.), Research and Exploration Co. 2011. Hydrology and water balance study, Karstic water resources and hard formation studies in Khalkhal area. Ardabil regional water company, Iran water resource Management Company. Department of Energy. Final report, vol. 4, 152p.

Peng, C.Y., Lee, K. and Ingersoll, G.M. (2002) An introduction to logistic regression analysis and reporting. Jour. Educat. Res., v.96 (1), pp.3-14.

Pohar, M., Blas, M. and Turk, S. (2004) Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study. Metodološki zvezki, v.1(1), pp.143-161.

Pradhan, B., Lee S., Mansor S., Buchroithner M.F. and Jallaluddin N. (2008) Utilization of optical remote sensing data and geographic information system tools for regional landslide hazard analysis by using binomial logistic regression model. Jour. Appld. Remote Sensing, v.2, pp.1-11.

Rautela, P. and Lakhera, R.C. (2000) Landslide risk analysis between Giri and Tons Rivers in Himachal Himalaya (India). Internat. Jour. Appld. Earth Observation and Geoinformation (JAG) 2 (3-4), pp.153–160.

Rupert, M.G., Cannon, S.H., Gartner, J.E., Michael, J.A. and Helsel, R.D., 2008. Using logistic regression to predict the probability of debris flows in areas burned by wildfires, Southern California, 2003-2006. USA: U.S. Department of the interior, U. S. geological Survey, Open-File Report, No. 2008-1370, 9p.

Saldivar-Sali, A. and Einstein, H.H. (2007) A Landslide risk Rating System for Baguio, Philippines. Engg. Geol., v.91 (2-4), pp.85-99.

Shariat, Jafari, M. (1996) Landslide, principles of natural slope stability (in Persian). Soil conservation and watershed management research institute, Tehran, Iran, 218p.

Shoaei, Z., Shoaei, G. and Emamjomeh, R. (2005) Interpretation of the mechanism of motion and suggestion of remedial measures using GPS continuous monitoring data *In:* K. Sassa, H. Fukuoka, Wang, F. and Wang, G., (Eds.), Landslides: Risk analysis and sustainable disaster management. Proc. first general assembly of the international consortium of landslides. Springer, pp.327-335.

Soeters, R. and van Westen, C.J. (1996) Slope instability recognition, analysis, and zonation. *In:* A.K. Turner and R.L. Schuster (Eds.), Landslides, Investigation and mitigation, Transportation research board special report 247. National Academy Press, Washington D. C., pp.129-177.

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. Science, v.240, pp.1285–1293.

Talaei, R., Ghayoumian, J., Shariat, Jafari, M. and Aliakbarzadeh, E. (2004) Study on effective factor causing landslide in southwest of Khalkhal region (in Persian). Agriculture research and education organization, Ministry of Jahad-e-Agriculture. Tehran, Iran, 153p.

Terlien, M.T.J., Van Asch, T.W.J. and Van Westen, C.J., (1995) Deterministic modelling in GIS-based landslide hazard assessment. *In:* A. Carrara and F.Guzzetti (Eds.) Geographical information systems in Assessing Natural Hazards. Kluwer, London, pp.57-77.

Thomas, G. (2003) Landslide occurrence as a response to land use change: a review of evidence from New Zealand. Catena, v.51(1-3), pp.297-314.

Turner, A.K. and Schuster, R.L. (1996) Landslides. Investigation and mitigation, Special Report 247, Transportation Research Board. National Academic, Washington, USA. 673p.

Uromeihy, A. and Mahdavifar, M.R. (2000) Landslide hazard zonation of the Khorshrostam area, Iran. Bull. Engg. Geol. Environ., v.58, pp.207-213.

Van Westen, C.J. (2000) The modeling of landslide hazard using GIS. Surveys in Geophys., v.21(2-3), pp.241-255.

Van Westen C.J., Seijmonsbergen, A.C. and Mantovani, F. (1999) Comparing landslide hazard maps. Natural Hazards, v.20, pp.137–158.

Van Westen, C.J. (1993) Application of geographic information system to landslide hazard zonation. ITC publication No. 15. International institute for aerospace survey and earth science (ITC), Enschede, Netherlands, 245p.

Varnes, D.J. (1984) Landslide hazard zonation: A review of principles and practice. IAEG Commission on landslides and other mass movements. Paris: UNESCO Press. 63p.

Wang, C., Esaki, T., Xie, M. and Qui, C. (2006) Landslide and debris-flow hazard analysis and prediction using GIS in Minamata-Hougawachi area, Japan. Environ. Geol., v.51, pp.91-102.

Wang, H.B. and Sassa, K. (2005) Comparative evaluation of landslide susceptibility in Minamata area, Japan. Environ. Geol., v.47, pp.956-966.

Wong, H.N., Lam, K.C. and Ho, K.K.S. (1998) Diagnostic report on the November 1993 natural terrain landslides on Lantau Island. GEO report no 69. The Government of the Hong Kong Special Administrative Region.

Wu, W. and Sidle, R.C. (1995) A distributed slope stability model for steep forested basins. Water Resource Res., v.31(8), pp.2097–2110.

Wu, S.R., Jin, Y.M., Zhang, Y.S., Shi, J.S., Dong, C., Lei, W.Z., Shi, L., Tan, C.X. and Hu, D.G. (2004) Investigations and assessment of the landslide hazards of Fengdu County in the reservoir region of the Three Gorges project on the Yangtze River. Environ. Geol., v.45, pp.560–566.

Yesilnacar, E. and Topal T. (2005) Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Engg. Geol., v.79 (3-4), pp.251-261.

Yin, K.L., and Yan, T.Z. (1988) Statistical prediction model for slope instability of metamorphosed rocks. *In:* C. Bonnard (Ed.), Proc. Fifth international symposium on landslides, Lausanne. A. A. Balkema, Rotterdam, Netherlands 2, pp.1269-1272.

Zweig, M.H. and Campbell G. (1993) Receiver Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. Clinical Chemistry, v.39(4), pp.561-577.

(Received: 1 November 2012; Revised form accepted: 21 March 2013)