A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate

Mehmet Lütfi Süzen · Vedat Doyuran

Abstract The purpose of this study is to evaluate and to compare the results of multivariate (logical regression) and bivariate (landslide susceptibility) methods in Geographical Information System (GIS) based landslide susceptibility assessment procedures. In order to achieve this goal the Asarsuyu catchment in NW Turkey was selected as a test zone because of its well-known landslide occurrences interfering with the E-5 highway mountain pass.
Two methods were applied to the test zone and two separate susceptibility maps were produced. Following this a two-fold comparison scheme was implemented. Both methods were compared by the Seed Cell Area Indexes (SCAI) and by the spatial locations of the resultant susceptibility pixels.
It was found that both of the methods converge in 80% of the area; however, the weighting algorithm in the bivariate technique (landslide susceptibility method) had some severe deficiencies, as the resultant hazard classes in overweighed areas did not converge with the factual landslide inventory map. The result of the multivariate technique (logical regression) was more sensitive to the different local features of the test zone and it resulted in more accurate and homogeneous susceptibility maps.

Keywords Landslide susceptibility · Multivariate · Bivariate · GIS · Asarsuyu · Turkey

Received: 15 May 2003 / Accepted: 17 September 2003
Published online: 18 December 2003

M. L. Süzen (✉) · V. Doyuran
Department of Geological Engineering,
Middle East Technical University, 06531 Ankara, Turkey
E-mail: suzen@metu.edu.tr
Fax: +90-312-2101263

Introduction

The unpredictable state of natural hazards constitutes one of the major uncontrolled impacts on the local and global economy. Almost no portion of the Earth's surface is free from the impact of natural hazards. In addition the continuous increase in population creates situations that are more vulnerable to geological hazards. This unwilling state of natural hazards is also supported by uncontrolled or poorly planned urbanization. Not very different from the majority of world, high precipitation, deficiencies in infrastructure and inadequacy of disaster plans and hazard maps are paid off by human lives in Turkey.
Over the past two decades, many scientists have attempted to assess landslide hazards and produced hazard/susceptibility maps portraying their spatial distribution. Two very general approaches can be listed as statistical and geomorphological (expert dependent) approaches (Soeters and van Westen 1996; Süzen 2002; Süzen and Doyuran 2004). Despite the methodological and operational differences, all methods proposed are founded upon a single basic conceptual model. This model requires first the identification and mapping of a set of geological-geomorphological factors which are directly or indirectly correlated with slope instability. Then it involves both an estimate of the relative contribution of these factors in generating slope failure, and ends with the classification of the land surface into zones of different hazard or susceptibility degrees (Soeters and van Westen 1996; Aleotti and Chowdhury 1999; Guzetti and others 1999; Süzen 2002).
Among the statistical methods, two significant groups arouse in the literature, the bivariate and the multivariate methods. In bivariate statistical analysis, each factor map is combined with the landslide distribution map, and weighting values based on landslide densities are calculated for each parameter class. Several statistical methods have been applied to calculate weighting values; these are: the landslide susceptibility method (Brabb 1984; van Westen 1992, 1993), information value method (Kobashi and Suzuki 1988; Yin and Yan 1988) and weights of evidence modeling method (Spiegelhalter 1986; Bonham-Carter 1996). On the other hand, in multivariate statistical analyses, the weights of causal factors controlling landslide occurrence indicate the relative contribution of each of these factors to the degree of hazard within a

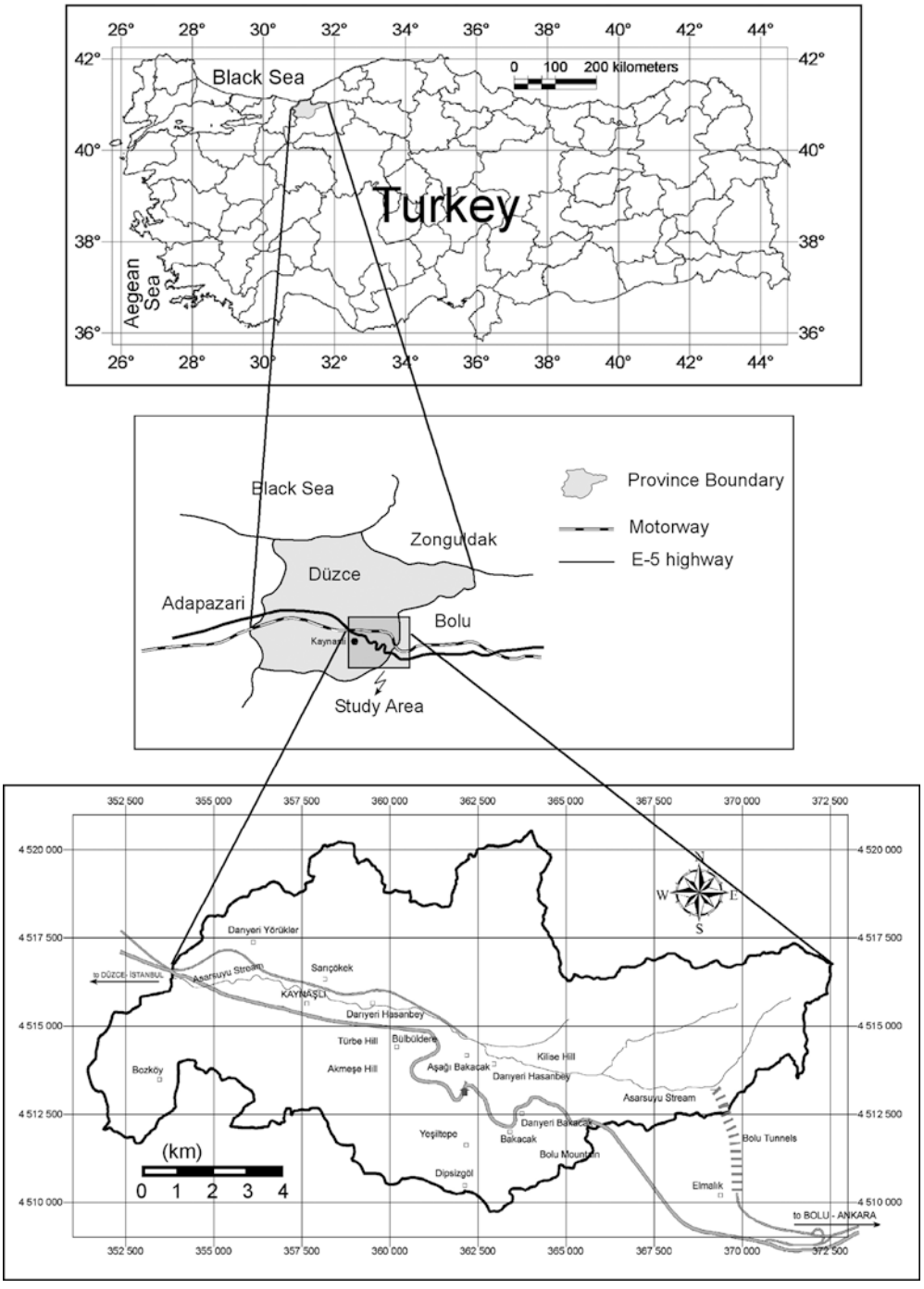


Fig. 1
Location map of the test area

defined land unit. Furthermore, the interactions between the factors are also encountered within the analyses. The common property of these analyses is their nature of being based on the presence or absence of stability phenomena within these previously defined land units (van Westen 1993). The first multivariate statistical analysis examples for landslide hazard zonation were published in Italy, mainly by Carrara (1983, 1988) and his colleagues (Carrara and others 1990, 1991, 1992). After their pioneering studies multivariate statistical studies flourished in the literature, which were mainly composed of applications of multiple regression and discriminant analyses. On the other hand, the use of binary logical regression, which is free of data distribution issues, are not so well exploited in the literature, only a few examples in the last few years can be observed, such as Atkinson and Massari (1998), Dai and others (2001) and Lee and Min (2001).

The aim of this study is to compare the bivariate (landslide susceptibility method) and multivariate (logical regression method) statistical landslide susceptibility assessment methods with respect to their accuracy and correctness in classification of susceptibility classes. In order to achieve these goals two methods were utilized in the Asarsuyu catchment which is selected as a test zone due to well-known landslide occurrences interfering with the Bolu Mountain E-5 highway pass (Fig. 1).

Data sets and database creation

Data sets

Three data sets were used to produce 13 different variables that were believed to control the landslide susceptibility in Asarsuyu catchment (Table 1). The data sets and the produced parameter maps are schematically shown in Fig. 2. These sets were based on remote sensing data (multispectral satellite imagery and aerial photographs), topographical maps, and existing geological maps. For all of the resultant parameter maps, raster spatial data model was used and the size of the pixels was chosen to be 25 by 25 m as the working scale was selected as 1:25,000. In all of the geo-spatial data related processes TNT MIPS of Microimages, an integrated GIS-RS program, was used. The remote sensing data set was composed of a part of Landsat 5 TM scene (Path 178, Row 32, dated 1994) and four sets of aerial photographs (1952, 1972, 1984 and 1994). The aerial photographs were used to create a multi temporal landslide inventory map regarding the activities in four periods. The photo characteristics of these slides were recorded simultaneously in the interpretation stage. This inventory map represents a training source set for further statistical analyses, in which a total of 49 landslides were distinguished (Fig. 3). In order to inquire the accuracy of the inventory map, locations of all the slides were verified in the field in the summer of 2000. Nearly all of the slides were observed in the field; however few ancient slides were obscured due to intense vegetation cover. The majority of the observed slides were of shallow translational or earth flow type. Only two landslides, the Bülbülderesi and Bakacak landslides (Fig. 3) have very large aerial extent and their morphology implies that they were formed by association of many smaller landslides (Süzen 2002). Therefore, based on their large extent and different mechanism (presence of gypsum layers yielded in siphons which directly feed the groundwater continuously), they were not included in the statistical analyses.

Table 1
The variables, their definitions and their ranges

Variable	Definition	Range (min-max) (unit)
LITHOMAP	Material properties	Categorical
DISTFAULT	Distance to fault line	0–4,791 (m)
FAULTDENS	Density of fault line in km^2	0–178 (#/km^2)
ELEVMAP	Elevation above mean sea level	220–1,580 (m)
DISTDRAIN	Distance to drainage lines	0–452 (m)
DRAINAGE DENS	Drainage density	12–352 (#/km^2)
DISTRIDGEMAP	Distance to ridges	0–658 (m)
ASPECTMAP	Aspect of the slopes	−1*–359 (degrees)
SLOPEMAP	Amount of slope	0–56 (degrees)
DISTSETTLEMENTMAP	Distance to settlement	0–6,093 (m)
DISTPOWER+ROAD	Distance to power lines and road network	0–2,312 (m)
DISTE-5MAP	Distance to E-5 highway	0–8,366 (m)
LANDCOVERMAP	Type of the land cover	Categorical

*−1 indicates flat areas

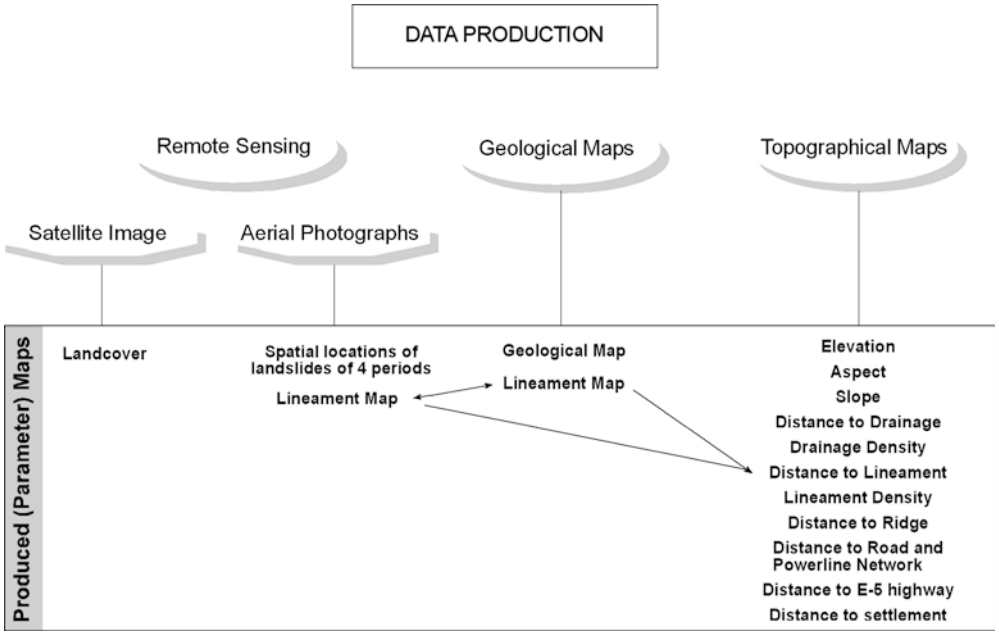


Fig. 2
The elements and products of the study in data production stage

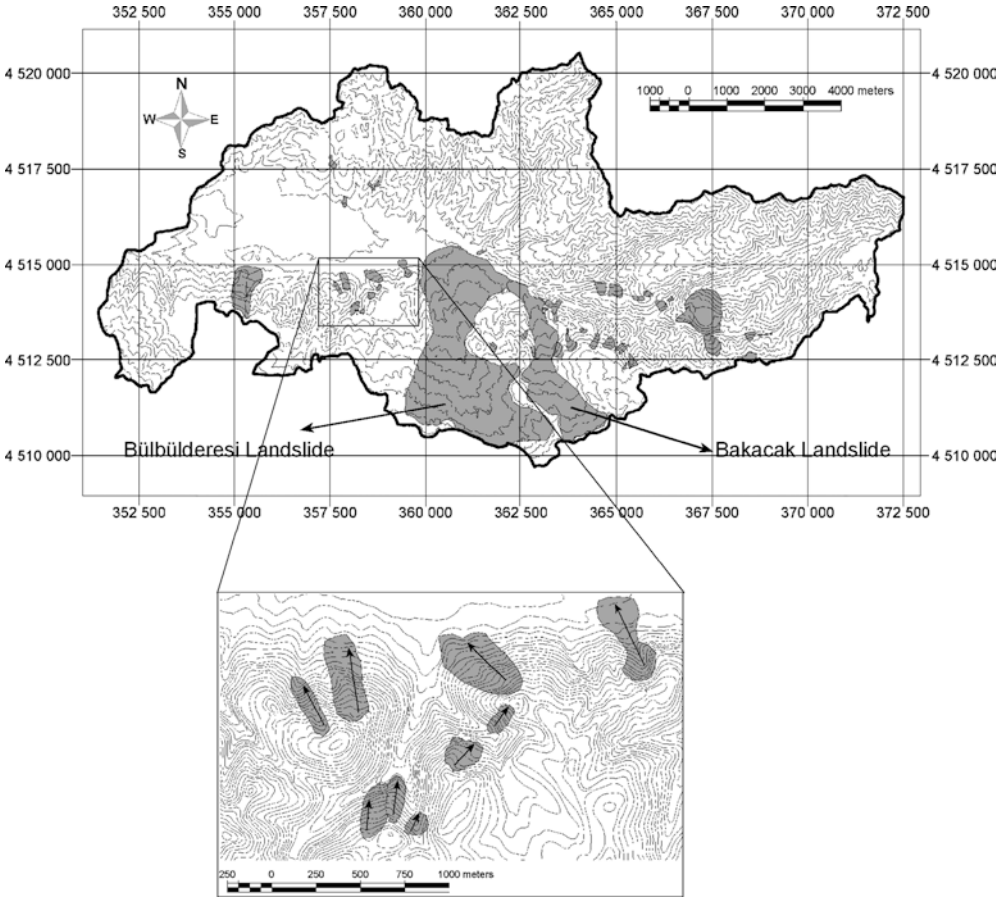


Fig. 3
Landslide inventory of Asarsuyu catchment

Among the remaining landslides, the areal extent of the smallest observable slide was approximately 2,000 m^2 and the largest was approximately 850,000 m^2.
Furthermore, during the interpretations of aerial photographs it was observed that landcover had changed drastically by extensive deforestation in 1970s and reforestation measures took place after the 1980s, and it was also observed that the number of landslides increased after deforestation (Süzen 2002). In order to include the effect of landcover, in the remote sensing data set, Landsat scene was used to extract the land cover information using various image processing and enhancement techniques (Süzen 2002, Süzen and Yeşilnacar 2002). Therefore, the produced landcover map was also treated as a parameter map in further statistical analyses.
The second data set was composed of geological data of the Asarsuyu catchment (Fig. 2). The geological map (Fig. 4) was compiled from the existing literature. Based on this compilation, it was seen that, the geology of the area is mainly composed of basement units such as Paleozoic metamorphics and intrusives, overlain by Mesozoic and Tertiary flyschoidal deposits and alluvium. The geological data set was also used to extract a lineament map, which was refined and verified by the help of aerial photographs and by field checks.
The third and the biggest data set was the topographical data set (Fig. 2). The source of this set was the 1:25,000 scale digital topographic maps of the study area. Eleven topographical parameters were extracted from these topographical maps regarding the morphology and the infrastructure information. The produced parameter maps are listed in Fig. 2. In converting the non-attributed discrete information into continuous maps, density of feature or true distance to feature algorithms were used (Süzen and Doyuran 2004).

Database production

In order to generate the decision rules of landslide occurrence a database concerning all of the parameters should be created. In order to reveal the landslide responsible pixels, a new approach, the so-called "seed-cells", was developed in the generation of decision rules of landslide occurrence. It was considered that the best-undisturbed morphological conditions (conditions before a landslide occurs) would be extracted from the close vicinity of the landslide polygon itself (Fig. 5). This was achieved by adding a buffer zone to the crown and flanks of the landslide. Following this a 25 m spaced grid was laid over the entire test area, having a point in the centers of each grid cell. The buffer zone was extracted from this vector grid file forming the seed cells (Fig. 5). These points were used to transfer the attributes of the thirteen parameter maps and used to create a single database referred to as the "seed cell" database, which will serve as a foundation for decision rule generation where the pre-slide conditions were simulated for further use in external statistical packages. A discussion on the "seed cell" concept and the methodology can be found in Süzen and Doyuran 2004.

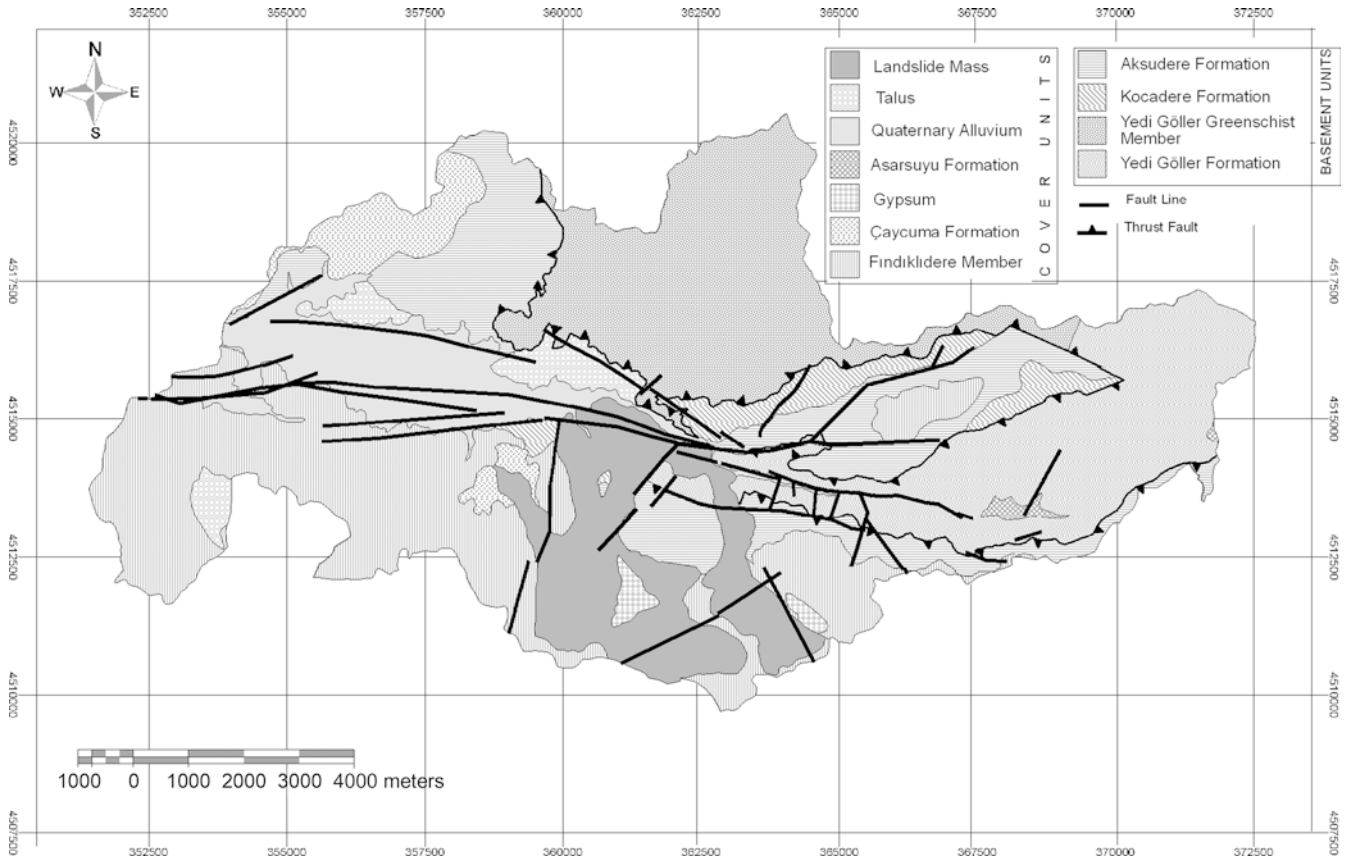


Fig. 4
The geological map of Asarsuyu catchment [modified from Erendil and others (1991)]

Landslide hazard assessment

In harmony with the aim of the study the two statistical methods were implemented to assess the landslide susceptibility in the Asarsuyu catchment. To achieve this goal, the nodes of seed cells were used as decision rule generators or training samples. Although there were 47 landslides (excluding the Bakacak and Bülbülderesi landslides) in the area, a small population for statistical analyses, the conversion of this inventory data into seed cells yielded in a number of 4,430 samples (seed cells) which is large enough to treat as a population. These 4,430 samples contain the attribute values of all of the thirteen parameters, and represented the pre-sliding conditions, which were known to result in slope instabilities.

Bivariate (landslide susceptibility) method

In bivariate analyses, the core of the analysis is to get the densities of landslide occurrences within each parameter map and within each parameter map's classes, and to get some data driven weights based on the class distribution and the landslide density. The brief outline of the method used is represented in Table 2.

In order to proceed with bivariate analysis, the continuous parameter maps would have to be converted into discrete (categorical) maps, to compute the responsible weights of each class. However, this conversion issue always remains unclear in landslide literature as most of the authors use their expert opinion for the boundaries of the classes (Carrara and Merenda 1974; Meneraud and Calvino 1976; Kienholz 1977, 1978; Stevenson 1977; Malgot and Mahr 1979; Ives and Messerli 1981; Kienholz and others 1983, 1988; Rupke and others 1988; Gupta and Joshi 1990; Pachauri and Pant1992; van Westen 1993; Soeters and van Westen 1996; Gupta and Anbalagan 1997). The use of expert opinion in defining the class boundaries may result in accurate explanation of the phenomenon and its mechanism in its test area, but it does not have an affinity to represent the same relation with the same data sets elsewhere. Based on this issue, a data driven methodology was proposed by Süzen and Doyuran (2004), as the classes should be selected according to the percentile divisions of seed cells. The data outside the seed cell range was discarded, as there were no landslides occurring in these regions. Further details of this conversion methodology, percentile map concept and the selection of class numbers were discussed in detail in Süzen and Doyuran (2004). In bivariate analyses, this data dependent division reduces the problem, of what weight should be given to each parameter map, as each class acts like a map and the ratio of the landslide density over the class area gives its natural weight.

After calculation of the weights, the weight values of the 13 parameters maps were added up to create the susceptibility map. No extra weighting procedure for the parameter maps were used in the summation process as the classes had been normalized and they received their natural weightings from the data itself.

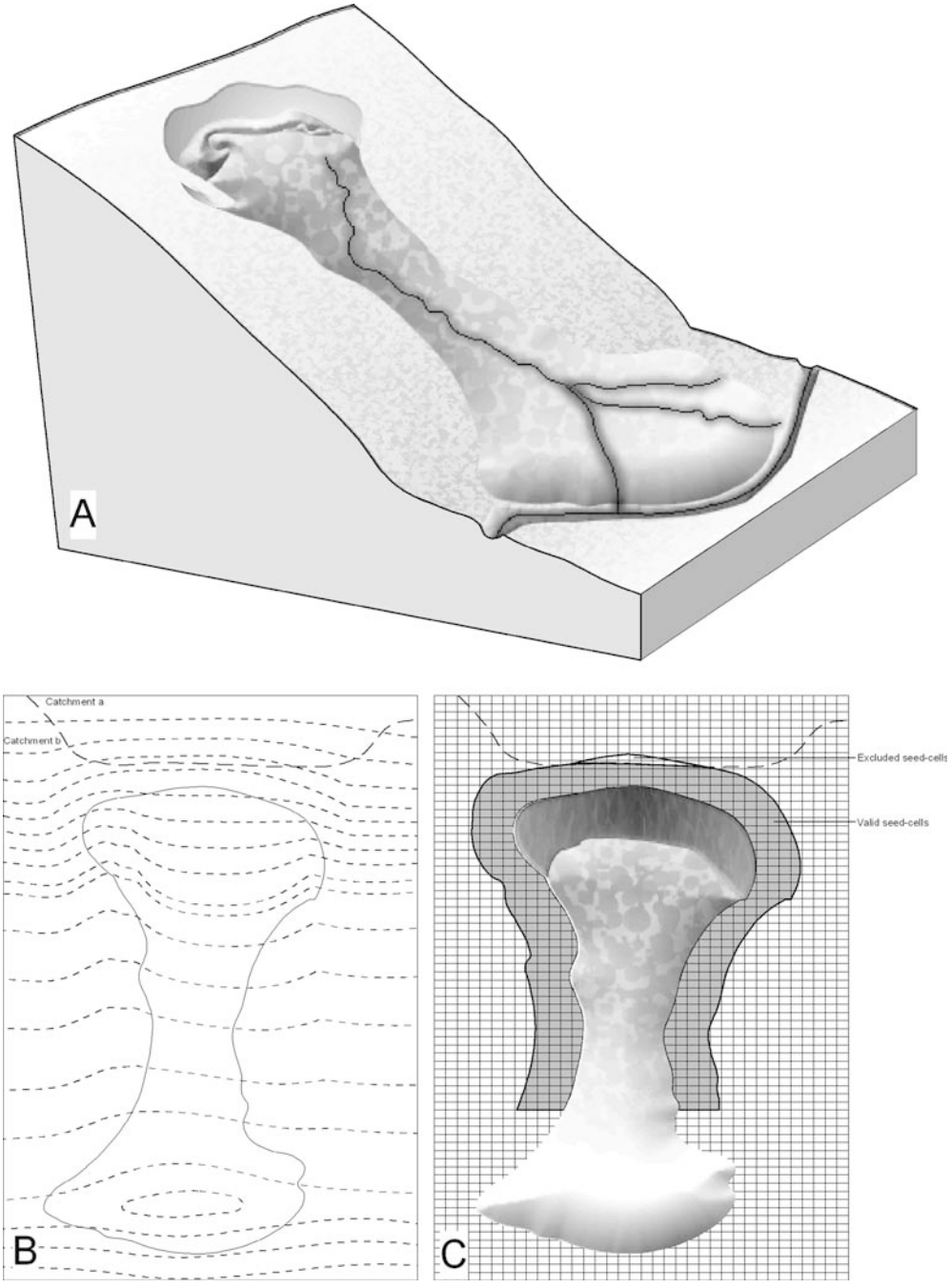
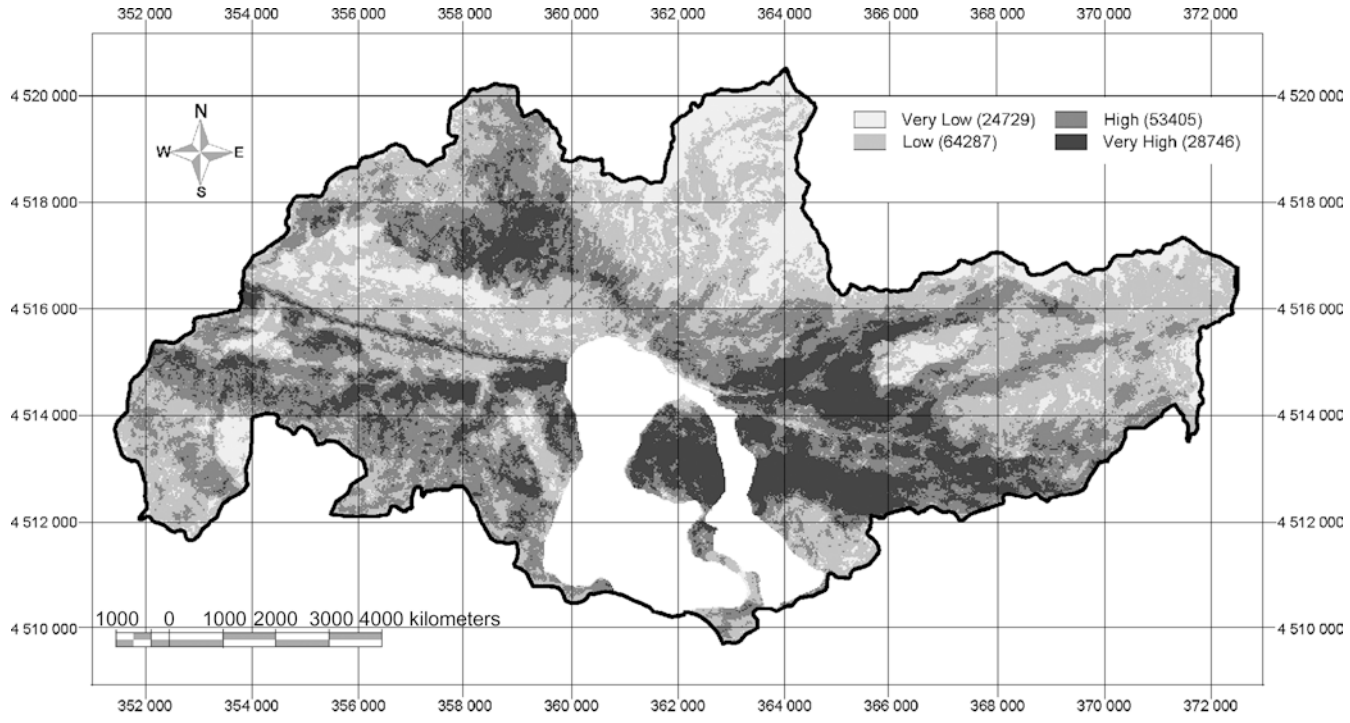


Fig. 5A–C
Graphical representation of "Seed-cells" concept. **A** Block diagram of the investigated landslide. **B** Map representation of the same slide with contour lines and the catchment boundaries. **C** Seed-cell buffer zone with excluded cells passing over the other catchment and the valid seed-cells. *Grey cells* show valid seed cells for analyses [Süzen and Doyuran (2004)]

Table 2
Outline of landslide susceptibility method

Bivariate analyses (landslide susceptibility)
$D_{\text{area}} = 1000\frac{Npix(SX_i)}{Npix(X_i)}$ In which D_{area} = areal density per millage, $N_{pix}(SX_i)$ = number of pixels with mass movements within variable class X_i, $N_{pix}(X_i)$ = number of pixels within variable class X_i. To evaluate the influence of each variable, weighting factors should have to be introduced, which compare the calculated density with the overall density in the area. The formula for the density-based area is: $W_{\text{area}} = 1000\frac{Npix(SX_i)}{Npix(X_i)} - 1000\frac{\sum Npix(SX_i)}{\sum Npix(X_i)}$

The resultant susceptibility map was then reclassified into 4 susceptibility zones (very low, low, high, very high) using the susceptibility maps distribution parameters. The mean value of the susceptibility map was taken as the pivot point and classes were assigned to the ± one standard deviations of the distribution. The resultant map and the landslide amounts in these susceptibility classes are given in Fig. 6. It was seen that 48% of the total area was classified as high



DN Value	Hazard class	% area covered	% landslide
0-93	Very low	14,45	0,5
93-146	low	37,56	6,2
146-199	High	31,20	26,7
199-481	Very High	16,79	66,6

Fig. 6
The susceptibility map and the amounts of landslides in each class as a result of bivariate analysis

and very high susceptibility class and within these classes 93.3% of seed cells were encountered. On the other hand 52% of the study area was classified as low and very low susceptibility, which in turn hosts only 6.7% of total seed cells in the area. This distribution also validates that the classification was quite reasonable.

Multivariate (logical regression) method

Logical regression allows forming a multivariate regression relation between a dependent variable and several independent variables that might affect the probability of the searched situation. If the searched variable is a dichotomous outcome the best method with free of predictor variable type seems to be logistic regression (Afifi and Clark 1998; Atkinson and Massari 1998; Dai and others 2001; Lee and Min 2001).

Binomial (or binary) logistic regression is a form of regression that is used when the dependent is a dichotomy and the independents are continuous variables, categorical variables, or both. The algorithm of logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable (the natural log of the odds of the dependent occurring or not). In this way, logistic regression estimates the probability of a certain event occurring (Atkinson and Massari 1998; Dai and others 2001; Lee and Min 2001).

The logistic model can be written in its simplest form as:

$$P = \frac{1}{1 + e^{-z}}$$

where P is the probability of an event occurring and also is the estimated probability of landslide occurrence. As z (linear logistic model) varies from $-\infty$ to $+\infty$, the probability varies from 0 to 1 on an s-shaped curve. And where z is defined as:

$$z = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \ldots\ldots\ldots\ldots + B_nX_n$$

where B_0 is the intercept of the model, n is the number of independent variables, $Bi(i = 1,2,3,\ldots..,n)$ is the slope coefficient of the model and $X_i(i = 1,2,3,\ldots..,n)$ is the independent variable.

In an extended form the equation of logistic regression could be written as:

Probability of belonging to population I

$$(\text{logit}) = \frac{1}{1+e^{B0+B1X1+B2X2+B3X3+\ldots\ldots\ldots\ldots+BnXn}}$$

Apart from the input data type flexibilities of logical regression, in general the advantage of logistic regression modeling over the other multivariate statistical techniques, including multiple regression analysis and discriminant analyses, is that the dependent variable can have only two values (a dichotomous outcome), and that predicted values can be interpreted as probability because they are

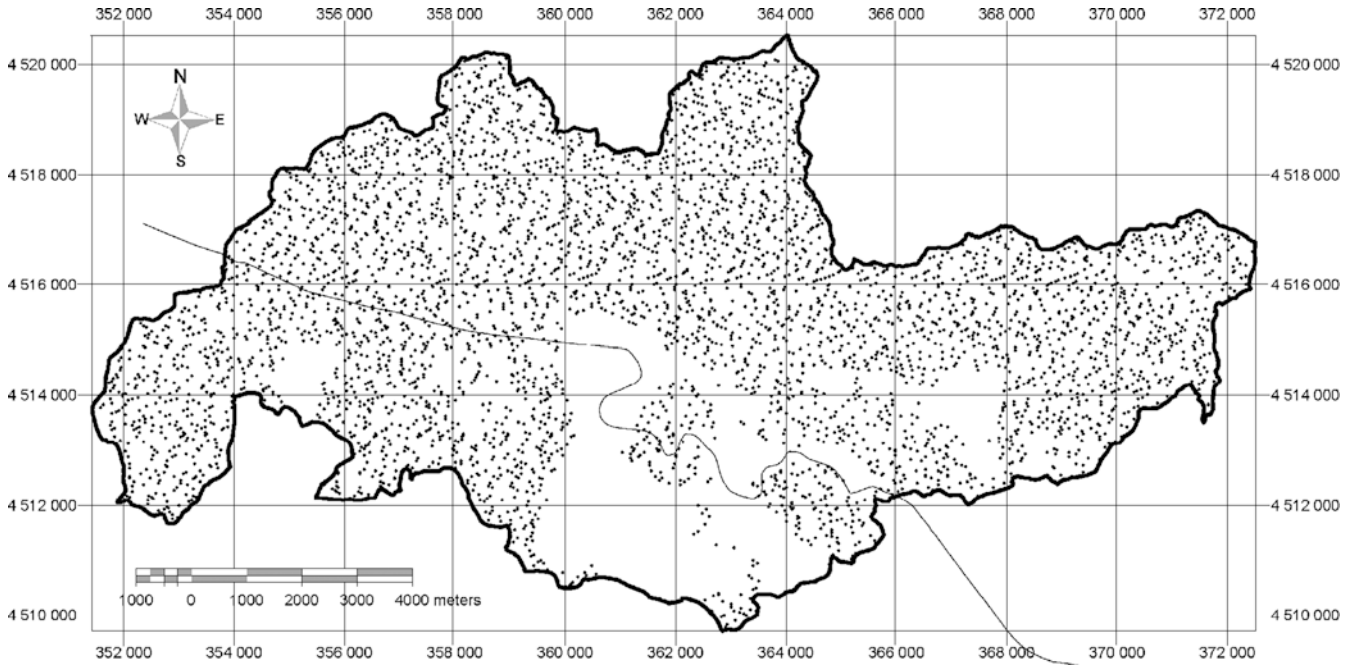


Fig. 7
The positions of selected 4430 random landslide free nodes

constrained to fall into an interval between 0 and 1 (Kleinbaum 1991). Logistic regression has many analogies to Ordinary Least Squares (OLS) regression: logit coefficients correspond to B coefficients in the logistic regression equation, the standardized logit coefficients correspond to beta weights, and a pseudo R^2 statistic is available to summarize the strength of the relationship. Unlike OLS regression, however, logistic regression does not assume linearity of relationship between the independent variables and the dependent, does not require normally distributed variables, does not assume homoscedasticity, and in general has less stringent requirements. The success of the logistic regression can be assessed by looking at the classification table, showing correct and incorrect classifications of the dichotomous, ordinal, or polytomous dependent. (Afifi and Clark 1998; Wrigley 1984).

However, in a strict sense, the logit function itself is not a probability function because the extrinsic parameters triggering the landslides such as the rainfall and earthquake vibration are not accounted for. It might be appropriate to term it as landslide susceptibility based on the intrinsic physical parameters.

In order to carry out the logical regression the total number of seed cells (4,430) was used. Moreover, another set of 4,430 random sample nodes were selected from the landslide free areas of Asarsuyu catchment and are presented in Fig. 7. Upon the selection of these random nodes, the values of the parameter maps were then transferred to the database of the random data set. Following the creation of the random data set database, the seed cells and the random set database were merged and a new column of a binary variable indicating the presence and absence of the landslides were added. This stage was repeated four times in order to have four different sets of random points, which in turn would let the user see if there was any convergence in the result of logistic regression analyses.

The initial assumption of the variables by the logistic regression is shown in Table 3. The system test reveals that the variables and the system constructed were valid. A Hosmer-Lemeshow test and Cox and Snell R square and Nagelkerke R square values were obtained in the range of 0.75 to 0.80, confirming that the statistical model supports the validity of the system with these variables.

After the validation of the system, logistic regression was applied to the data set and the resultant classification table is presented in Table 4. The system overall has a success of classifying 77.3% of the pixels correctly. The 77.3% was the highest classification success among the four different random sets, though the remaining three were oscillating around 75% with plus minus 1%.

Based on the loadings of the variables after logistic regression the logistic regression equation was compiled as follows:

Table 3
The initial assumption of the variables. Landbin refers to landslide binary as 0 values indicate no landslide and 1 indicates presence of landslide[a,b]

				Predicted		
				Landbin		Percentage correct
				0	1	
Step 0	Observed	Landbin	0	0	4,430	0
			1	0	4,430	100.0
		Overall percentage				50.0

[a]Constant is included in the model,
[b]the cut value is 0.500

$$Z = 0.773046364 + (0.130082590 * \text{GEOCODE}) - (0.004154044 * \text{DRAINAGE DENSITY}) - (0.000897442 * \text{DISTANCE TO POWERLINES AND ROAD NETWORK}) - (0.004813297 * \text{DISTANCE TO RIDGE}) + (0.000212306 * \text{DISTANCE TO FAULT}) - (0.000525944 * \text{DENSITY OF FAULT}) - (0.00114028 + \text{DISTANCE TO SETTLEMENT}) (0.001257937 * \text{DISTANCE TO E} - 5 \text{ HIGHWAY}) - (0.001047155 * \text{DISTANCE TO DRAINAGE}) - (0.120726833 * \text{LANDCOVER}) - (0.002760724 * \text{ELEVATION}) + (0.52982916 * \text{SLOPE}) - (0.00296956 * \text{ASPECT})$$

The observed groups and the predicted probabilities of these groups are presented in Fig. 8. It was clearly seen that, in the values larger than the cutoff value, were under dominance of binary variable 1 which represents slope instability and the opposite of this argument was valid also for the values lower than the cut off value as safe (landslide free) pixels. In the cut off value it was seen that both probabilities were nearly the same, with a little emphasis on the landslides side.

The resultant susceptibility map was then reclassified into four susceptibility zones (very low, low, high, very high) using the scores of logistic regression. The class boundaries were as follows: 0–0.25 very low susceptibility, 0.25–0.5 low susceptibility, 0.5–0.75 high susceptibility and 0.75–1 very high susceptibility. This re-classification of the susceptibility map is shown as the landslide susceptibility map of Asarsuyu catchment (Fig. 9).

When the shapes of the susceptibility zones were analyzed, the susceptibility map produced from logical regression results in more homogenous zones than that of bivariate analyses (Figs. 6, and 9), especially in the end members of the zones; in very low and very high susceptibility classes. The low and very low susceptibility classes constitute 72.31% of the area with corresponding 22.23% of the total landslide seed cells. On the other hand, the rest of the area was classified as high and very high susceptibility that yield in 27.68% of the area with corresponding 77.77% of the total landslide seed cells (Fig. 9).

Comparison of susceptibility maps

Two susceptibility maps were produced from bivariate analysis and logical regression analyses. Both of them produced acceptable results, as both classify the majority of the seed cells in high or very high susceptibility classes. The produced susceptibility maps were then tested in the field by the slope instabilities which occurred after the 12 November 1999 Kaynasli earthquake (M_s = 7.2). Approximately 20 new or reactivated landslides were observed in the field and all of these slides fall into high and very high classes of the produced susceptibility map. However as the shapes and areas of susceptibility classes of the two maps were differing, they have to be analyzed in order to reveal which method was more successful and which method was more accurate. Therefore, two comparison schemes were implanted.

Comparison of methods via their areas and corresponding landslide seed cells

It was shown that both methods classify less than half of the study area as high or very high susceptibility, while two thirds of the seed cells were located in these classes, which could be said to be a success. However, when the class areas were normalized with the landslide seed cell counts some important issues have arisen. In order to normalize the areas, the area percent values were divided with the landslide seed cell percent values. The resultant values were called the seed cell area index (SCAI), which shows the density of landslides among the classes (Table 5).

The logic behind SCAI lies in the correct classification of seed cells within a very conservative areal extent. As a result, it was desired that the high and very high susceptibility classes should have very small SCAI values and low and very low susceptibility classes to have higher SCAI values.

When the SCAI values of the two methods were compared it was found that the multivariate susceptibility map (MSM) produced from logical regression analyses had lower SCAI values than that of bivariate version. Only in the very low susceptibility class did the bivariate susceptibility map (BSM) have a better result. In low and high susceptibility classes MSM had a clear superiority,

Table 4
The final classification of logistic regression. See Table 3 for definition of Landbin[a]

				Predicted		
				Landbin		Percentage correct
				0	1	
Step 1	Observed	Landbin	0	3,387	1,019	76.9
			1	985	3,445	77.8
		Overall percentage				77.3

[a]The cut value is 0.500

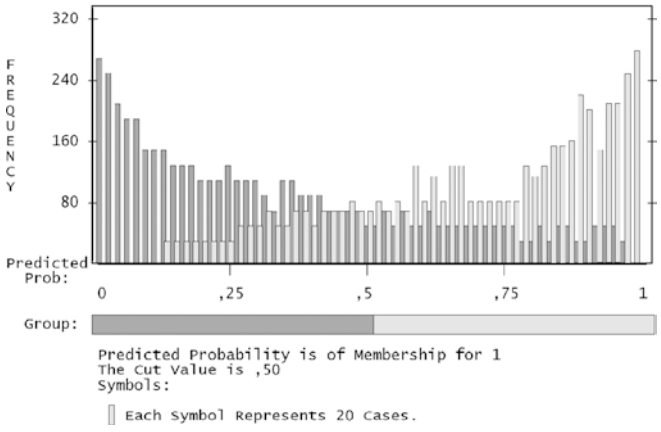


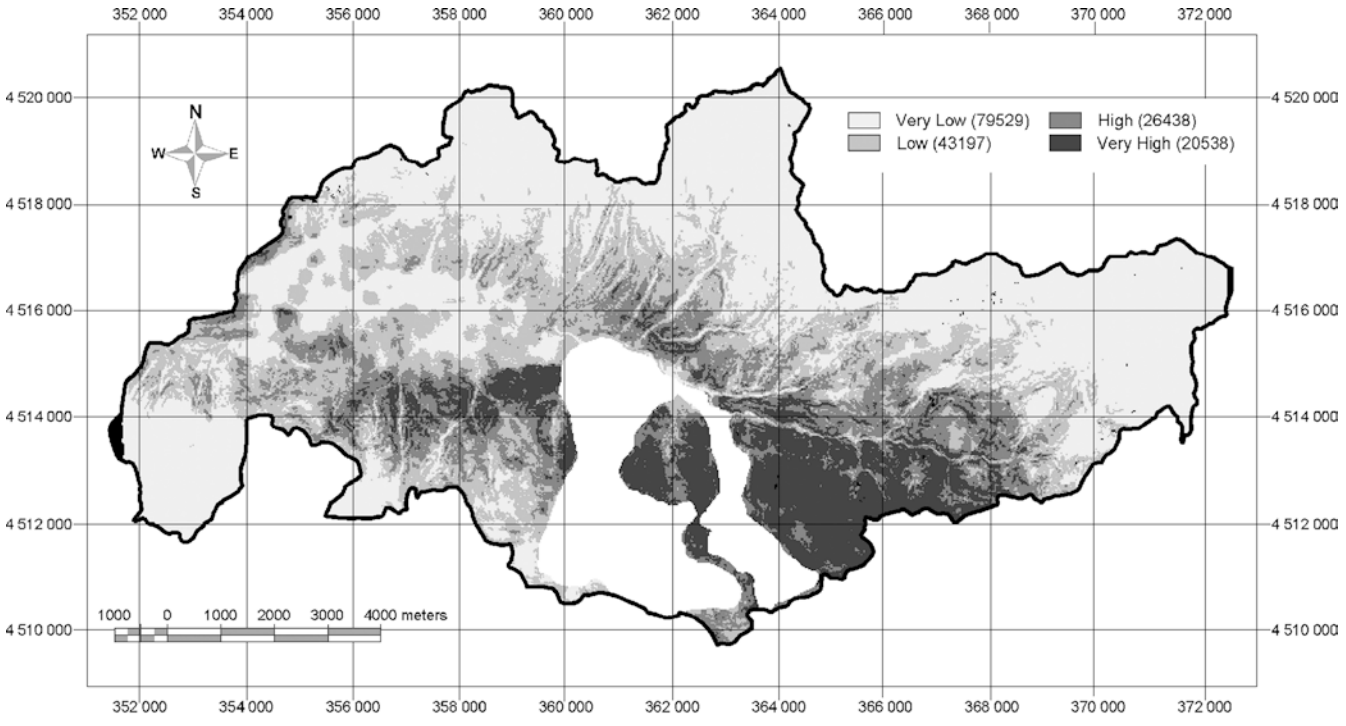
Fig. 8
Observed groups and predicted probabilities (*0*=No landslide, *1*=Landslide)

however in very high susceptibility class they were quite close to each other but the logical one had a slight advantage as less area was classified as very high susceptibility. Although the BSM had a high SCAI value, which was desirable for the very low susceptibility class, the area classified was only 14.45% of the total area, which was very limited for urban planning purposes. The system should be a little more flexible rather than a mechanically rigid system, considering the acceptable risk of the dwellers in the area. From Table 5 the 3.47% increase of landslide seed cells in the very low susceptibility class in the MSM ended up in 32.41% of extra area with minimal susceptibility, which could be accepted.

The comparison of two methods in the spatial domain

The SCAI does not reveal any information about the change of the susceptibility score within a pixel with respect to a change in the method. In order to achieve the pixel basis changes or mismatches, both of the susceptibility maps were first re-classified into known numerical values. The bivariate susceptibility map was reclassified as 1, 2, 3, 4 starting from very low susceptibility and ending up with very high susceptibility, correspondingly. The logical regression susceptibility map was reclassified by values 10 times the class numbers of the bivariate susceptibility map, accordingly the class values were 10, 20, 30, and 40. After this re-coding process the two maps were added up. The possible outcomes were collated in a reclassification table, that is presented in Table 6.
As can be seen in Table 6, some combinations result in misclassified pixels. These were dominated by the absence of a susceptibility score in both or in one of the

Fig. 9
The hazard map and the amounts of landslides in each class as a result of multivariate analysis



DN Values	Hazard class	% area covered	% landslide
0-0,25	Very low	46,86	3,97
0,25-0,5	low	25,45	18,26
0,5-0,75	High	15,58	29,48
0,75-1	Very High	12,10	48,28

Table 5
The densities of landslides among hazard classes of both methods

		Area (%)	Seed (%)	SCAI
Bivariate (BSM)	Very low	14.45	0.5	28.8945
	Low	37.56	6.4	5.8684
	High	31.20	26.8	1.1642
	Very high	16.79	66.8	0.2514
Logical (MSM)	Very low	46.86	3.97	11.79
	Low	25.45	18.26	1.3939
	High	15.58	29.48	0.5284
	Very high	12.10	48.28	0.2506

Table 6
The available combinations of re-coding process and their meanings

BIVARIATE (BSM) \ LOGICAL (MSM)	0	10	20	30	40
1	11	21	31	41	
2	12	22	32	42	
3	13	23	33	43	
4	14	24	34	44	

Misclassification
Correct Classification
Acceptable Classification
Not Acceptable Classification

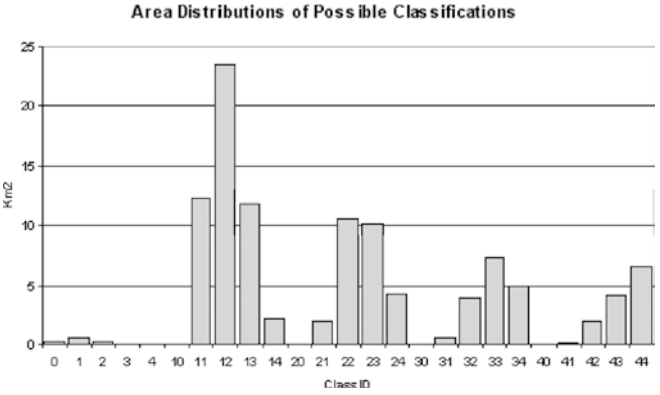


Fig. 10
The areal distributions of re-classified pixels

susceptibility maps, which results from a small acceptable resampling error in the order of a few meters of deviation in central pixel coordinates especially in the map edges. The amount of these pixels compared to the whole classified area was merely 1.12%, which was negligible (the sum of 0, 1, 2, 3, 4, 10, 20, 30 and 40 class ID's in Fig. 10). If the two susceptibility maps converge in the same susceptibility classes after summation, as such they possess 11, 22, 33 and 44 susceptibility ID's and they were called correctly classified pixels. The areal extent of these pixels constituted 36.67 km^2 and 34.16% of the total area (Fig. 10). The locations of the correctly classified pixels are shown in Fig. 11.

Another pixel association was observed as the acceptable classification when the susceptibility classes in both of the susceptibility maps differed from each other by one rank in the susceptibility classification scheme. For example, a change from a very low susceptibility to a low susceptibility, or change from a high susceptibility to a very high susceptibility was acceptable. This association was indicated in the summation map by the following ID's 12, 21, 23, 32, 34 and 43. The area covered with this association was 48.69 km^2 corresponding to 45.36% of the total area. The locations of acceptable pixels are shown in Fig. 12.

The unacceptable pixels were defined as the difference of susceptibility ranks having more than one rank. For example, a class rank change from a very low susceptibility to a very high susceptibility was not acceptable. These pixels were represented in the summation as 13, 14, 24, 31, 41 and 42. The area covered by these not acceptable pixels was 20.79 km^2 with corresponding 19.6% of the area. The locations of these pixels are shown in Fig. 13.

Upon the investigation of the not accepted pixels of the two methods, it was seen that the bivariate susceptibility map was overestimating the susceptibility classes relative to the logical susceptibility map. As shown in the legend of Fig. 13, six pixel values were present and the first three of them had greater occurrences than the remaining three. The first three were 13, 14 and 24, indicating that they belong to the high and very high susceptibility class in the bivariate map and the low to very low susceptibility class in the logistic regression map. Although this was a relative comparison, it could be said that MSM is underestimating the susceptibility; however, the classification scheme fits well in the remaining three pixel values. If MSM was underestimating the susceptibility, the remaining three pixel counts should be more than that of the observed values. A further investigation was made in order to find the reasons why BSM overestimates the susceptibility with the aid of initial parameter maps, percentile maps and the susceptibility maps. It was found that most of the errors were dependent on the class boundaries of the parameter maps and on the unsegmented nature of linear features. For example, the first percentile of distance from E-5 had the most weight between all percentiles, which was the result of the faint E-5 highway trace (A in Fig. 13) in the western part of Asarsuyu catchment. However it is known that only the mountain pass of the highway had landslides, the remaining valley part of the highway is free of landslides. Also the fault density and distance to fault percentile maps were responsible from the mismatch of the two susceptibility maps. Furthermore, area B in Fig. 13 coincides with peculiar lithological units having fault related boundaries. In these pointed areas these units do not have landslides but they appear to be high to very high susceptibility class as they gain high weight values from other outcrops of these lithological units. On the other hand, the good correlation in the high and very high susceptibility classes of both methods should not be underestimated.

In order to see where both susceptibility methods have the same and acceptable classifications, the acceptable pixels and the correct pixels were added up, which represents an acceptable classification of nearly 80% of the area. This result is presented in Fig. 14.

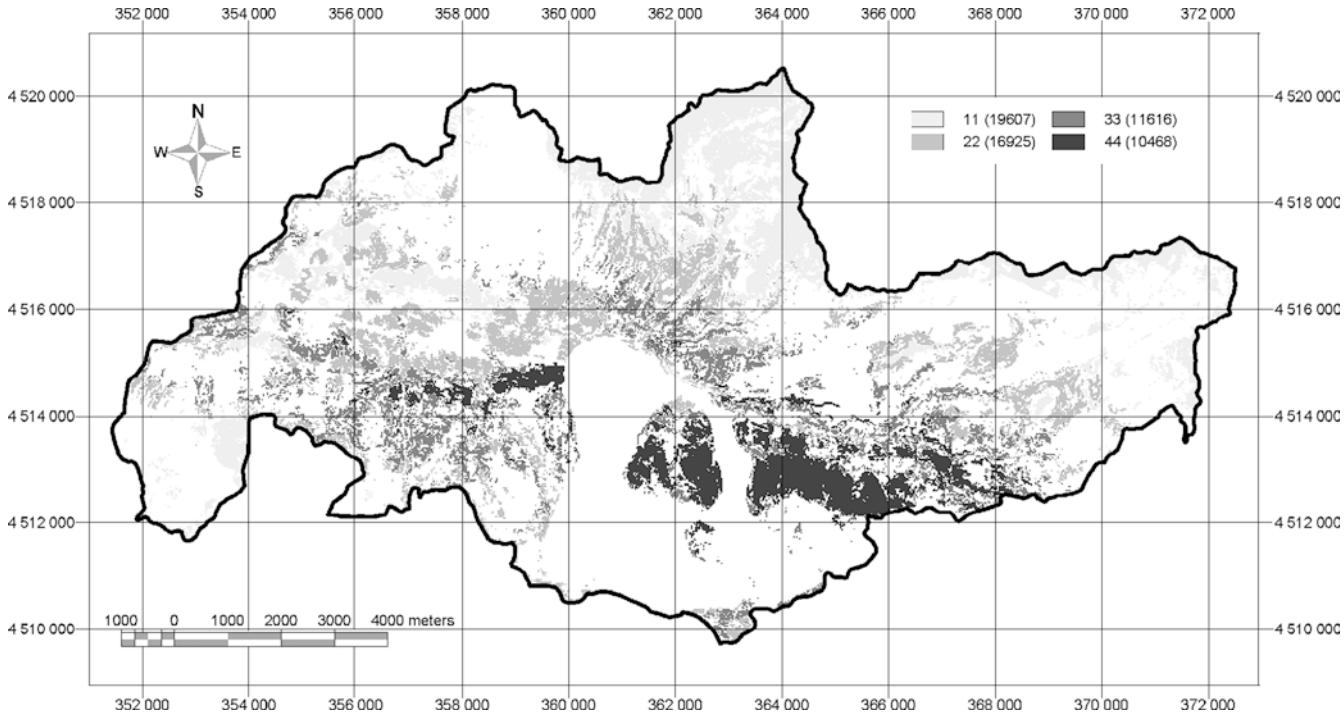
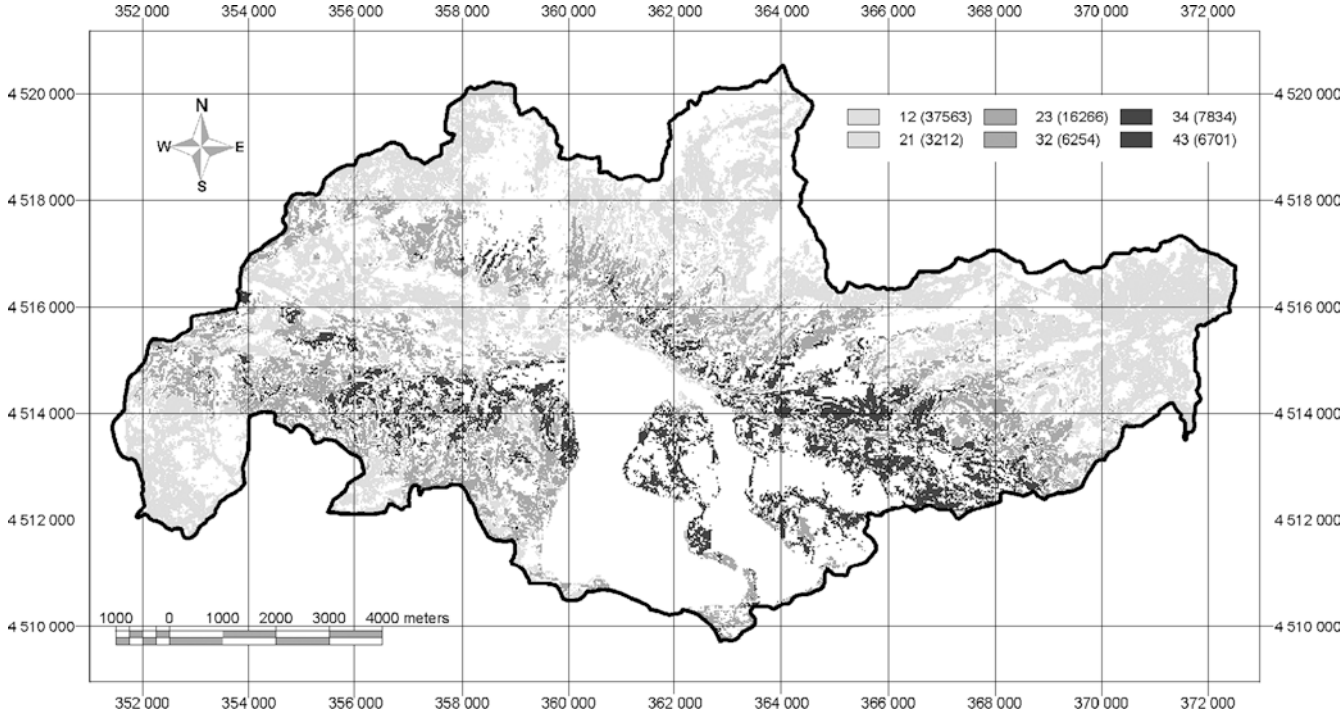


Fig. 11
The locations of the correctly classified pixels (*numbers* in the legend indicate the pixel counts)

Discussions and conclusions

Upon the completion of the susceptibility maps, a quantitative comparison scheme was implemented. In order to do this, a two-fold methodology was followed. The first was to compare the areas of susceptibility classes and the corresponding densities of landslides. This analysis was quite valuable as the idea in optimum zonation refers to allocating minimum areas for high susceptibility zones, while covering most of the landslides present in the area. Based on this optimum zonation concept, an index was defined as the Seed Cell Area Index (SCAI) and the two susceptibility maps were compared. The logical regression susceptibility map was found to be more accurate and possesses

Fig. 12
The locations of the acceptable pixels (*numbers* in the legend indicate the pixel counts)



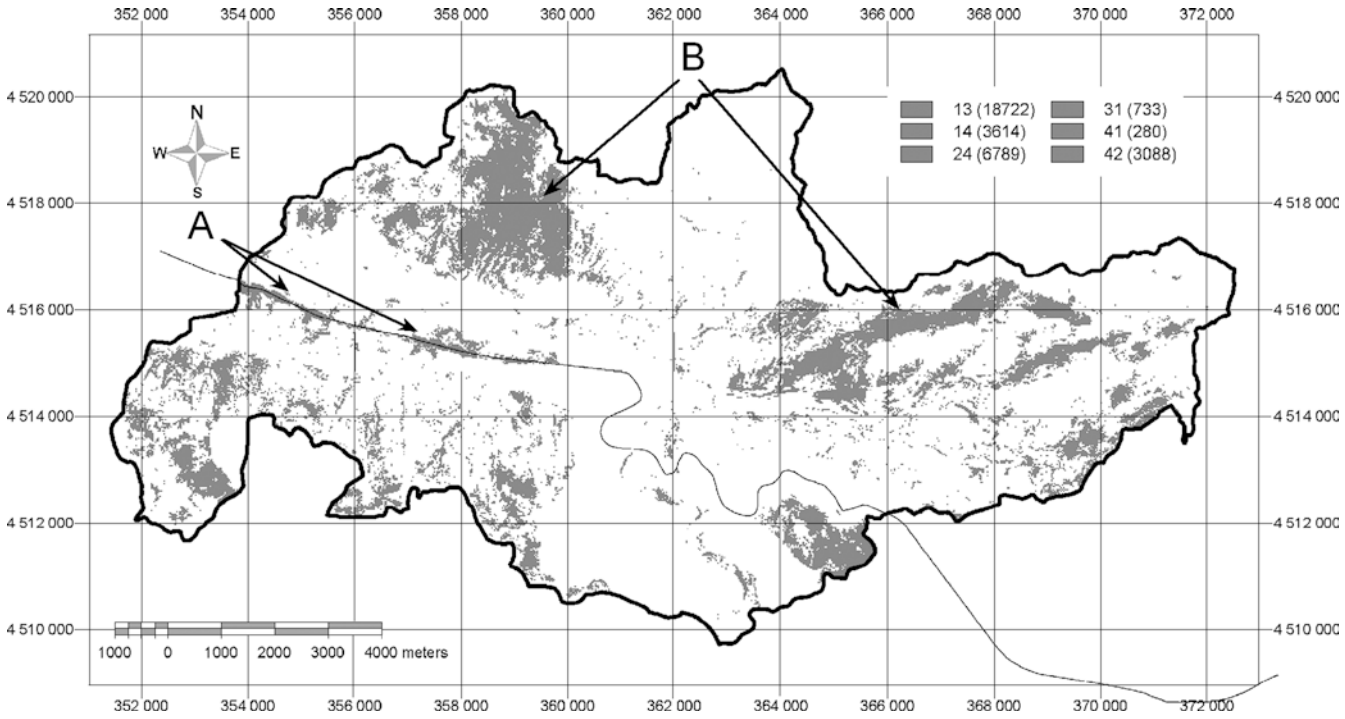
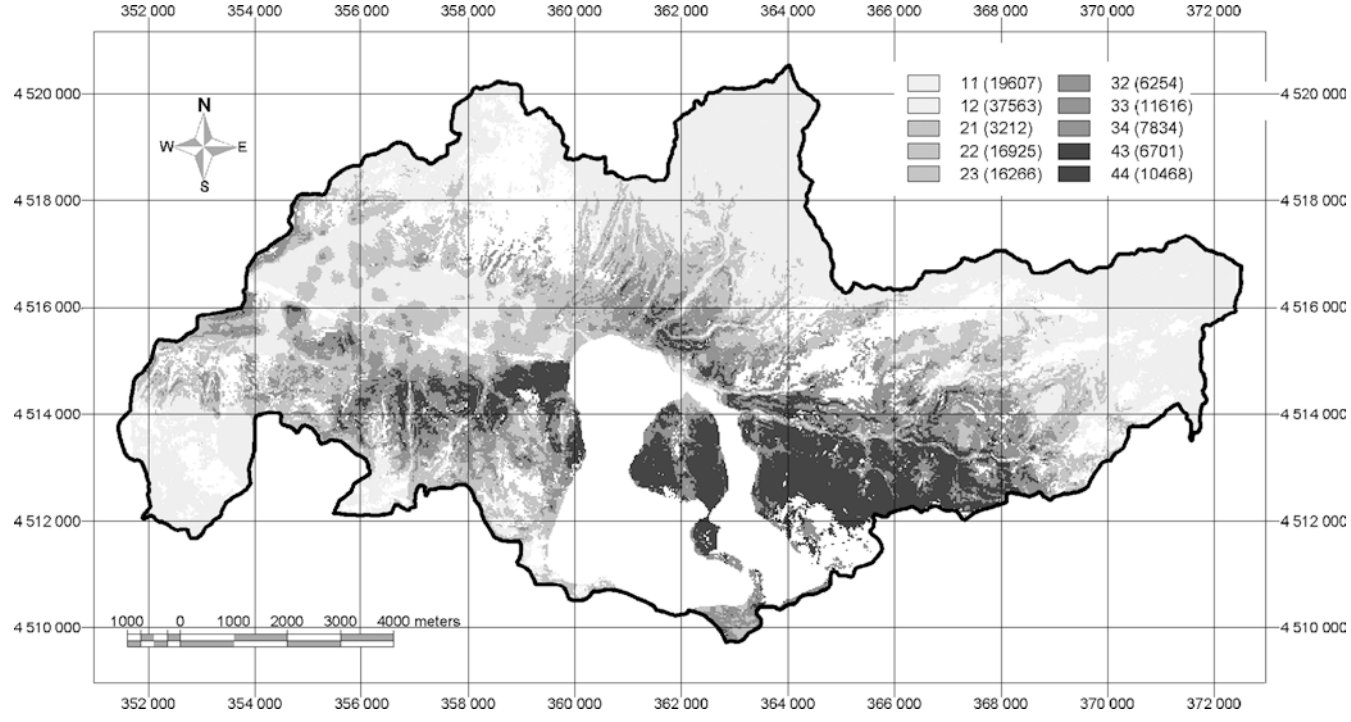


Fig. 13
The locations of the not acceptable pixels (*numbers* in the legend indicate the pixel counts)

more acceptable results relative to that of bivariate analyses.

The second method was to compare the susceptibility classes of the two produced maps via their spatial locations. In order to achieve this, a re-coded matrix was prepared. It was found that 80% of the two maps were converging into acceptable results. The remaining mismatched 20% of the area was reflected by the deficiencies of bivariate analyses. The overweighing of percentiles of distance to E-5 highway, distance to fault, fault density and the geological map were the major sources of this mismatch. Further subdivisions in these parameter maps should be done in order to increase the sensitivity of the bivariate analyses.

Fig. 14
The locations of the correctly classified and the acceptable pixels united (*numbers* in the legend indicate the pixel counts)



The resultant susceptibility map of logical regression analysis seems to be more consistent in the spatial locations of similar class ranks; such few pixels of very high risk were surrounded with low to very low susceptibility class pixels. In BSM they have more disseminated appearances. Therefore a hole-filling algorithm altering the susceptibility classes seems to be indispensable.

Based on these arguments about the differences of susceptibility mapping methodologies, the speed of bivariate methods could be said to be an advantage over multivariate methods. However, the necessity of being more accurate and more sensitive to the modeled feature, considering the intermingling of parameters, places multivariate methods as superior to bivariate methods. The necessity to have a statistical background should be cited as a potential limiting factor for extensive usage of multivariate methods for susceptibility mapping.

After it had been decided to use the logical regression susceptibility map as the final output, it was seen that, the northern slope of Asarsuyu catchment was definitely classified as a very low landslide susceptibility area. The attributes responsible for this classification were quite reasonable as these areas were the least populated, the land cover was not disturbed and the cover was dense forest, very distant to the E-5 highway and to the major active fault, the lithology was resistant enough, although the drainage density and slope values were higher than the rest of the area.

On the other hand the southeastern slopes belong definitely to the very high susceptibility class. The reasons could be listed as the removal of lateral supports by E-5 highway cut slopes, close location to active faults, high disturbance to land cover, heavy traffic resulting in extra vibration and the presence of flyschoidal units.

Acknowledgements This study was financially supported by the Middle East Technical University Research Project No: AFP-99–03–09–04.

References

Afifi AA, Clark V (1998) Computer aided multivariate analysis. Chapman Hall, London, 455 pp

Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. Bull Eng Geol Env 58:21–44

Atkinson PM, Massari R (1998) Generalized linear modeling of susceptibility to landsliding in the central Apennines, Italy. Computers and Geosciences 24(4):373–385

Bonham-Carter GF (1996) Geographic information systems for geoscientists, modeling with GIS. Pergamon Press, Oxford, 398 pp

Brabb EE (1984) Innovative approaches to landslide hazard and risk mapping. In: Proc 4th Int Symp on Landslides. Canadian Geotechnical Society, Toronto, vol 1, pp 307–324

Carrara A (1983) Multivariate models for landslide hazard evaluation. Mathematical Geology 15(3):403–427

Carrara A (1988) Landslide hazard mapping by statistical methods: a "Black Box" approach. In: Workshop on natural disaster in European Mediterranean countries, Perugia, Italy. Consiglio Nazionale delle Ricerche, Perugia, pp 205–224

Carrara A, Merenda L (1974) Metodologia per un censimento degli eventi franosi in Calabria. Geologia Applicata e Idrogeologia 10:237–255

Carrara A, Cardinalli M, Detti R, Guzzetti F, Pasqui V, Reichenbach P (1990) Geographical information systems 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. Universita degli Studi de Milano, pp 17–28

Carrara A, Cardinalli M, Detti R, Guzzetti F, Pasqui V, Reichenbach P (1991) GIS techniques and statistical models in evaluating landslide hazard. Earth Surf Proc Land 16(5):427–445

Carrara A, Cardinali M, Guzzetti F (1992) Uncertainity in assessing landslide hazard and risk. ITC J 2:172–183

Dai FC, Lee CF, Xu ZW (2001) Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. Environ Geol 40(3):381–391

Erendil M, Aksay S, Kuşçu İ, Oral A, Tunay G, Temren A (1991) Bolu masifi ve çevresinin jeolojisi. MTA Rap No: 9425 (unpublished)

Gupta P, Anbalagan R (1997) Slope stability of Tehri dam reservoir area, India using landslide hazard zonation (LHZ) mapping. Q J Eng Geol 30:27–36

Gupta RP, Joshi BC (1990) Landslide hazard zoning using the GIS approach – A case study from the Ramganga catchment, Himalayas. Eng Geol 28:119–131

Guzzetti F, Carrara A, Cardinalli M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-case study, central Italy. Geomorphology 31:181–216

Ives JD, Messerli B (1981) Mountain hazard mapping in Nepal: introduction to mountain research project. Mountain Research and Development 1(3–4):223–230

Kienholz H (1977) Kombinierte geomorphologische gefahrenkarte 1:10,000 von Grindelwald. Catena 3:265–294

Kienholz H (1978) Maps of geomorphology and natural hazards of Grindelwald, Switzerland, scale 1:10,000. Arctic Alpine Res 10:169–184

Kienholz H, Bichsel M, Grunder M, Mool P (1983) Kathmandu Kakani area, Nepal: mountain hazards and slope stability map. United Nations University, Mountain Hazards mapping project, map 4, Scale 1:10,000

Kienholz H, Mani P, Klay M (1988) Rigi Nordlene, Beurteilung der naturgefahren und waldbaliche prioritatenfestlegung. In: Proc INTRAPRAEVENT, Graz, Austria, vol 1, pp 161–174

Kleinbaum DG (1991) Logistic regression a self-learning text. Springer, Berlin Heidelberg New York

Kobashi S, Suzuki M (1988) Hazard index for the judgment of slope stability in the Rokko Mountain region. In: Proc INTERPRAEVENT, 1988, Graz, Austria, vol 1, pp 223–233

Lee S, Min K (2001) Statistical analysis of landslide susceptibility at Yongin, Korea. Environ Geol 40:1095–1113

Malgot J, Mahr T (1979) Engineering geological mapping of the west Carpathian landslide areas. Bull Int Assoc Eng Geologists 19:116–121

Meneraud JP, Calvino A (1976) Carte ZERMOS, zones exposees a des risques lies aux mouvements du sol et du sous-sol a 1:25,000, region de la Moyenne Vesubie (Alpes Maritimes). Bureau de Recherches Geologiques Minieres, Orleans, France, 11 pp

Pachauri AK, Pant M (1992) Landslide hazard mapping based on geological attributes. Eng Geol 32:81–100

Rupke J, Cammeraat E, Seijmonsbergen AC, van Westen CJ (1988) Engineering geomorphology of the Widentobel catchment, Apenzell and Sankt Gallen, Switzerland: a geomorphological

inventory system applied to geotechnical appraisal of slope stability. Eng Geol 26:33–68
Soeters R, van Westen CJ (1996) Slope instability recognition analysis and zonation. In: Turner KT, Schuster RL (eds) Landslides: investigation and mitigation. Transportation Research Board National Research Council, Special Report No 247, Washington, DC, pp 129–177
Spiegelhalter DJ (1986) Uncertainty in expert systems, in artificial intelligence and statistics. Addison Wessey, Reading, MA, pp 17–55
Stevenson PC (1977) An empirical method for the evaluation of relative landslide risk. Bull Int Assoc Eng Geol 16:69–72
Süzen ML (2002) Data driven landslide hazard assessment using geographical information systems and remote sensing, M.E.T.U. PhD Thesis, Unpublished, 196 pp. http://www.metu.edu.tr/~suzen
Süzen ML, Doyuran V (2004) Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. Eng Geol (in press)
Süzen ML, Yeşilnacar E (2002) A landcover extraction scheme for landslide hazard assessment using satellite image components. Int Conf Environ Problems of the Mediterranean Region, Near East University, Nicosia, North Cyprus, 11–16 April, (in press)
Van Westen CJ (1992) Medium scale landslide hazard analysis using a PC based GIS: a case study from Chinchina. In: Alzate JB (ed) Proc 1st Simposio Internacional sobre Sensores Remotes y Sistemas de Informacion Geografica (SIG) para el Estudo de Riesgos Naturales, Bogota, Colombia. Instituto Geografico Agustin Codazzi, Bogota, vol 2, 20 pp
Van Westen CJ (1993) Application of geographic information systems to landslide hazard zonation. ITC Publication No 15, International Institute for Aerospace and Earth Resources Survey, Enschede, The Netherlands, 245 pp
Wrigley N (1984) Categorical data analysis for geographers and environmental scientist. Longman, Harlow
Yin KL, Yan TZ (1988) Statistical prediction model for slope instability of metamorphosed rocks. In: Bonnard C (ed) Proc 5th Int Symp Landslides, Lausanne. Balkema, Rotterdam, vol 2, pp 1269–1272