Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Engineering Geology 71 (2004) 303–321
ENGINEERING GEOLOGY
www.elsevier.com/locate/enggeo

# Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey

Mehmet Lütfi Süzen*, Vedat Doyuran

*Remote Sensing and Geo. Inf. Sys. Lab., Department of Geological Engineering, Middle East Technical University, 06531 Ankara, Turkey*

Received 25 June 2002; accepted 25 April 2003

**Abstract**

In the last decades, landslide hazard assessment has attracted many researchers' attention. A number of parameters are suggested to be responsible to quantitatively explain the mechanism of landslides; many of these parameters are very important and factual. However, some data types and models are site-specific and could not be applied to different locations. Furthermore, the data stored in continuous parameter maps are divided into a number of classes arbitrarily, depending on the vision of the expert. Basically, this division controls the result of bivariate analysis. Besides, the responsible portion of the parameter map controlling the mechanism is also weighted arbitrarily. Based on these two facts, the class boundaries put a prejudice on the produced susceptibility/hazard maps, which result in dependence on the knowledge of the user rather than being dependent on the data and the fact itself. The aim of this study is to refine the previously defined methods in a more data-dependent trend. To achieve this goal, two new concepts: seed cells and percentile maps are introduced. Seed cells are the zones that are considered to represent the best undisturbed morphological decision rules (conditions before landslide occurs) and would be achieved by adding a buffer zone to the crown and flank areas of the landslide. To quantitatively classify the input parameter maps, the data distributions of seed cells in the parameter maps are divided into a number of classes on the basis of their distribution's percentile break-points upon which the parameter maps are directly dependent on the seed cell distributions, hence to the data itself.


*Keywords:* Landslide susceptibility mapping; Geographical information systems; Asarsuyu; Turkey

## 1. Introduction

Every year, thousands of people all over the world are losing their lives in natural disasters, which have large impacts on the local and global economy. Almost no portion of the earth's surface is free from the impact of natural hazards. Due to increasing population densities and uncontrolled or not well-planned development, more and more people are affected by disasters. Not very different from the majority of the world, high precipitation, deficiencies in infrastructure and lack 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 susceptibility maps portraying their spatial distribu-

* Corresponding author. Fax: +90-312-210-1263.
*E-mail address:* suzen@metu.edu.tr (M.L. Süzen).

0013-7952/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0013-7952(03)00143-1

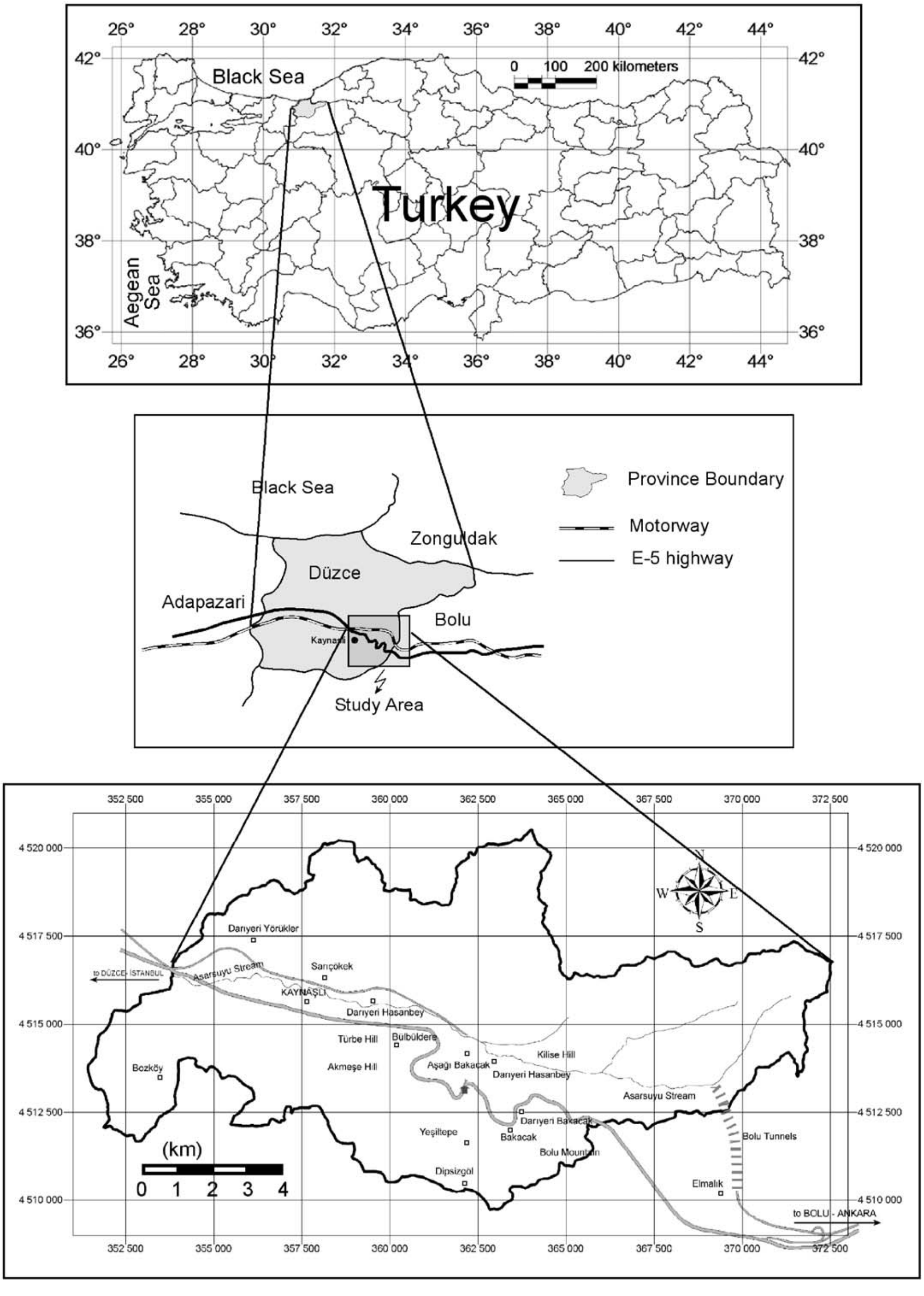



Fig. 1. The geographical location of the study area.

tion. However, up till now, there has been no general agreement on the methods or even on the scope of these investigations. Despite the methodological and operational differences, all methods proposed are based upon a single basic conceptual model. This model requires first the identification and mapping of a set of geological–geomorphological factors that are directly or indirectly correlated with slope instability. Then, it involves both an estimate of the relative contribution of these factors in generating a slope failure, and the classification of the land surface into zones of different susceptibility degrees (Süzen, 2002).

Several hazard modeling methodologies have been presented in the literature and could be aggregated into two groups: the heuristic and the statistical approaches (Soeters and van Westen, 1996). In the heuristic approach, all of the decision rules to create the hazard or susceptibility map are given by the researcher and they directly depend on the skill of the researcher, so expert knowledge is indispensable. These methods generally consist of application of geomorphological rules to the landslide-prone areas and they flourished the literature in the past two decades. Since they depend on the skill of the map producer, their reproducibility by other researchers with the same data set seems to be not viable, which incorporates a subjective state to the produced maps. Some good examples of the heuristic-type studies can be listed as Kienholz (1978), Malgot and Mahr (1979), Ives and Messerli (1981) and Rupke et al. (1988). In spite of their accurate and proficient maps, the degree of subjectivity and expert dependency in these methods lead the scientists to mine around the available statistical methods, which gave thrust to the use of statistical methods in the last decade.

All forms of statistical methods are used to estimate the relative contributions of the factors responsible for slope instability, and to create some predictions based on these factors; however, they still have some differences in the way that they handle the data. They can be categorized into two subgroups according to their data analysis methods as: bivariate and multivariate methods (Soeters and van Westen, 1996; Süzen, 2002). In multivariate methods, all of the factors are treated together and their interactions assist to resolve the phenomena statistically; whereas bivariate methods assume that the factors are not correlated with each other. There have been many applications of bivariate methods in the literature, some of which can be listed chronologically as: Brabb (1984), Kobashi and Suzuki (1988), Yin and Yan (1988), Gupta and Joshi (1990), Pachauri and Pant (1992), Van Westen (1993), Fiener and Haji (1999) and Uromeihy and Mahdavifar (2000).

The aim of this study is to refine the previously defined bivariate statistical landslide susceptibility assessment methods in a more data-dependent approach. To achieve this goal, two new ideas ("seed-cells" for defining the decision rules of slope instabilities and percentile class divisions for transforming the continuous variables into categorical variables) are introduced in this study. The Asarsuyu catchment is selected as a test zone due to well-known landslide occurrences interfering with Bolu Mountain E-5 highway pass and the Anatolian Motorway (Fig. 1).

## 2. Input data and data production

In harmony with the purpose of the project, three data sets were used to generate the necessary variables (Fig. 2). These sets are remote sensing products (multispectral satellite imagery and aerial photographs), topographical maps, and existing geological maps. The size of the pixels for all of the products is chosen as $25 \times 25$ m as the working scale is selected as 1:25,000. In all of the geo-spatial data-related processes, TNT MIPS of Microimages, an integrated GIS-RS program, is used.

The first step through any susceptibility assessment scheme is the evaluation of the current situation, preferably having information about the past states of the hazard. In this research, the spatial locations of the landslides in four different periods (1952, 1972, 1984 and 1994) were extracted from the aerial photographs and transformed into base maps. Simultaneously, the interpretation is acquainted with a photo checklist regarding the observable characteristics of the landslides. The landslides of these four different periods are merged into one vector file forming the cumulative landslide inventory map of the catchment. This vector file represents a training source for statistical analyses to figure out the parameters controlling the instability. A total of 49 landslides is distinguished in four periods in the Asarsuyu catchment. To inquire the accuracy of the inventory map, locations of all the

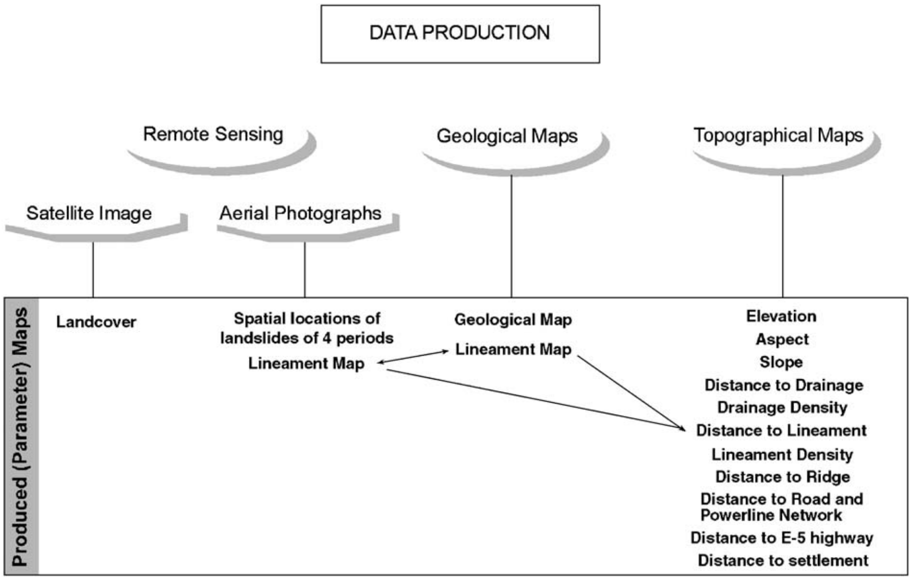



Fig. 2. The elements and products of the study in data production stage.

slides are verified in the field in the summer of 2000. Nearly all of the slides are observed in the field; however, some ancient slides were obscured due to intense vegetation cover. Majority of the observed slides are of shallow earth flow or translational. Only two landslides, the Bülbülderesi and Bakacak landslides (Fig. 3), have very large aerial extents and their morphology implies that they are produced by association of many smaller landslides. Therefore, based on their large extent and different mechanism, they are not included in the statistical analyses. Among the remaining landslides, the areal extent of the smallest observable slide is approximately 2000 $m^2$ and the largest is 850,000 $m^2$ approximately.

Apart from the aerial photographs, in the remote sensing data set, Landsat TM 5 (Path 178/Row 32, dated 1994), is used to extract the land cover information using various image processing and enhancement techniques (Süzen, 2002; Süzen and Yeşilnacar, 2002). The previous aerial photo interpretations show that landcover is changed drastically by extensive deforestation in the 1970s and reforestation measures took place after the 1980s, and it is also observed that the number of landslides increased after deforestation (Süzen, 2002). Therefore, the produced landcover map is also treated as a parameter map in further statistical analyses.

The second data set is the geological map data set. This data set is based on compilation of the various scaled geological maps produced by the General Directorate of Mineral Research and Exploration (Ketin, 1955; Batum, 1968; Orkan et al., 1977; Aktimur et al., 1983; Erendil et al., 1991; Gedik and Alkaş, 1996; Sözen et al., 1996) and existing literature (Blumenthal, 1948; Aydın et al., 1987; Cerit, 1990; Dalgıç, 1994; Işın, 1999). Upon this compilation, the geology of the area is mainly composed of Paleozoic metamorphic intrusive basement and its covering Mesozoic and Tertiary flyschoidal deposits. However, these maps have been prepared for general geological purposes without considering the special needs of the susceptibility assessment procedures, so the geological units are regrouped according to lithological characteristics rather than to their stratigraphic content and age. The geological data set is also used to extract a lineament map, which is refined and verified by the help of aerial photographs and by field checks.

The third and the largest data set are the 1:25,000 scale digital topographical maps obtained from General Command of Mapping (Turkish Army). Eleven spatial and topographical parameters were extracted from these topographical maps regarding the morphology and the infrastructure information (Fig. 2). The information present in these maps consists of

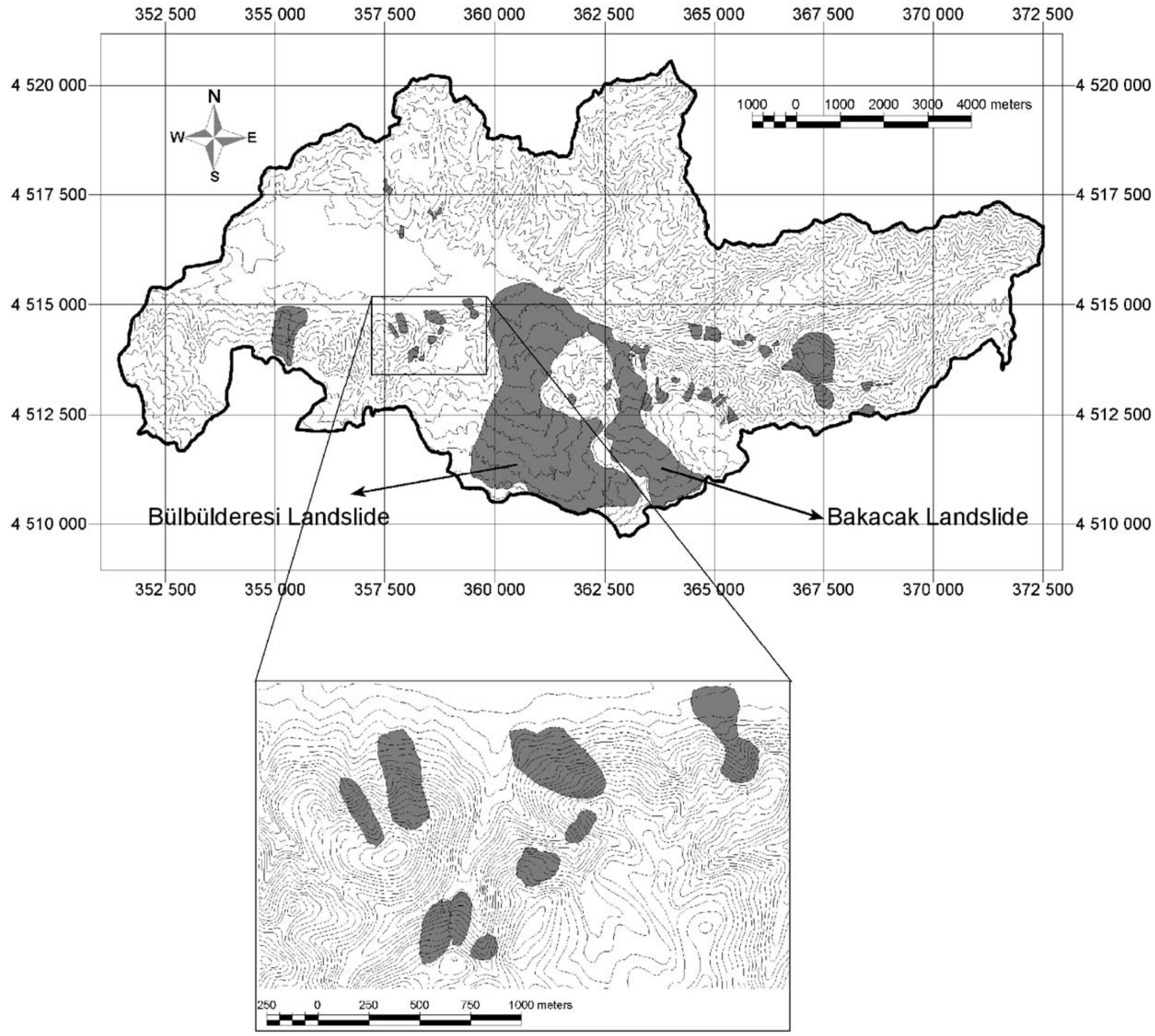



Fig. 3. Landslide inventory of Asarsuyu catchment.

vector coverages as represented by point, line or polygonal features. Although this discrete form is not the most suitable way to be used in a GIS, yet it has to be converted into continuous surfaces. This spatial transformation from discrete objects to continuous surfaces is controlled by the type of the attribute that the objects possess. When the vector data has elevation (*Z*) attributes (e.g., topographical contours), this transformation is achieved through kernel estimators via interpolation (Bonham-Carter, 1996). The produced surfaces from attribute-bearing vectors in this project include elevation, aspect and slope maps. On the other hand, if the vector object has no *Z* attributes (e.g., infrastructure information), which is quantified by the presence or absence of the spatial object, a non-interpolative method should have to be implemented. This non-interpolative method is either the density (number of line/point elements of fixed length in a fixed area) of the object within a specified area, or the nearest distance of the pixels to that object. For the non-attributed vectors such as drainage, fault, watershed and infrastructure vectors, both

Table 1
The used parameters and their data ranges

| Variable | Definition | Range (min–max) (unit) |
|---|---|---|
| LITHOMAP | Material properties | Categorical- |
| DISTFAULT | Distance to fault line | 0–4791 (m) |
| FAULTDENS | Density of fault line in kilometer squares | 0–178 (#/km$^2$) |
| ELEVMAP | Elevation above mean sea level | 220–1580 (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[a]–359 (degrees) |
| SLOPEMAP | Amount of slope | 0–56 (degrees) |
| DISTSETTLEMENTMAP | Distance to settlement | 0–6093 (m) |
| DISTPOWER + ROAD | Distance to power lines and road network | 0–2312 (m) |
| DISTE-5MAP | Distance to E-5 highway | 0–8366 (m) |
| LANDCOVERMAP | Type of the land cover | Categorical |

[a] − 1 indicates flat areas.

distance and density functions are used in this study to yield the related continuous surface maps (Table 1). Moreover, for maps requiring distance functions, in this study, an innovative approach is considered by calculating the true distances to the features rather than taking the horizontal map distances. Six maps out of 13 are corrected topographically; they are distance to drainage, distance to lineament, distance to ridge, distance to E-5 highway, distance to power and road network and distance to settlement. The reason for using such a correction is to include the effect of topographic gradient in the generated decision rules. It is obvious that implementation of this topographical gradient correction has more effect on steep slopes,

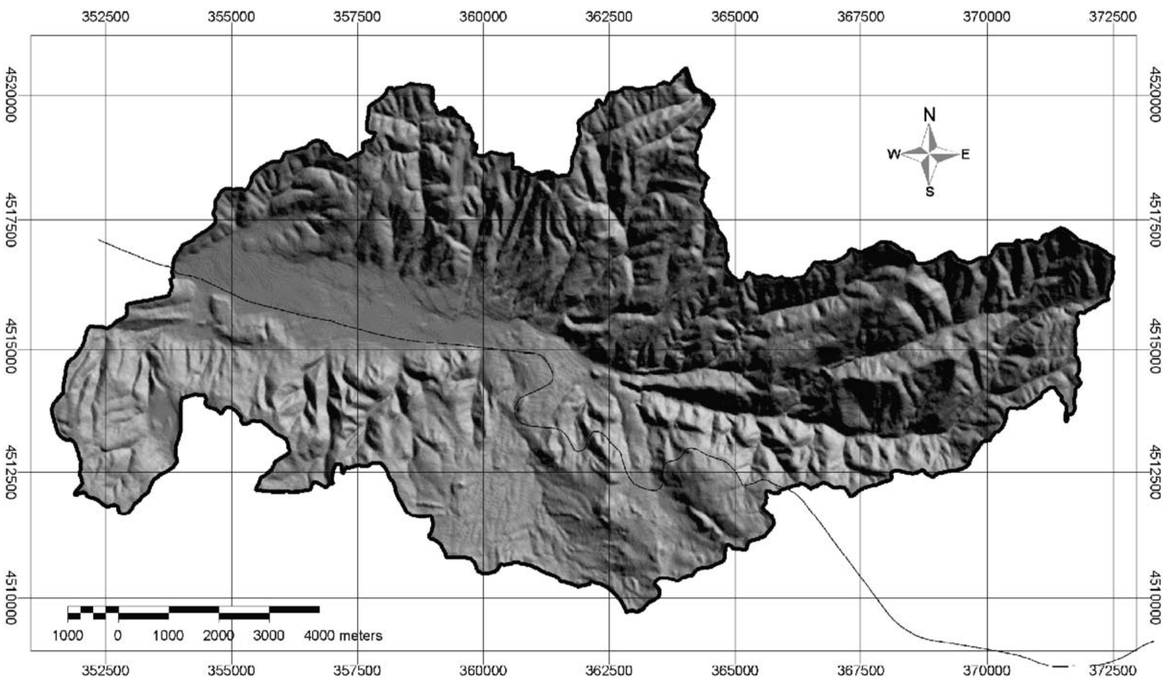



Fig. 4. A sample of parameter maps—relief shaded digital elevation model of the Asarsuyu catchment.

because in gentle slope or in flat land, the two distances would be nearly equal.

To achieve standardization through the three domains, all of the maps were resampled using a reference raster as the digital elevation model (DEM). This resampling is vital to fix the extents and the centre of the pixels of each parameter map.

The parameter maps produced include the information of a particular parameter considering the whole of the study area; the relief (elevation) map is

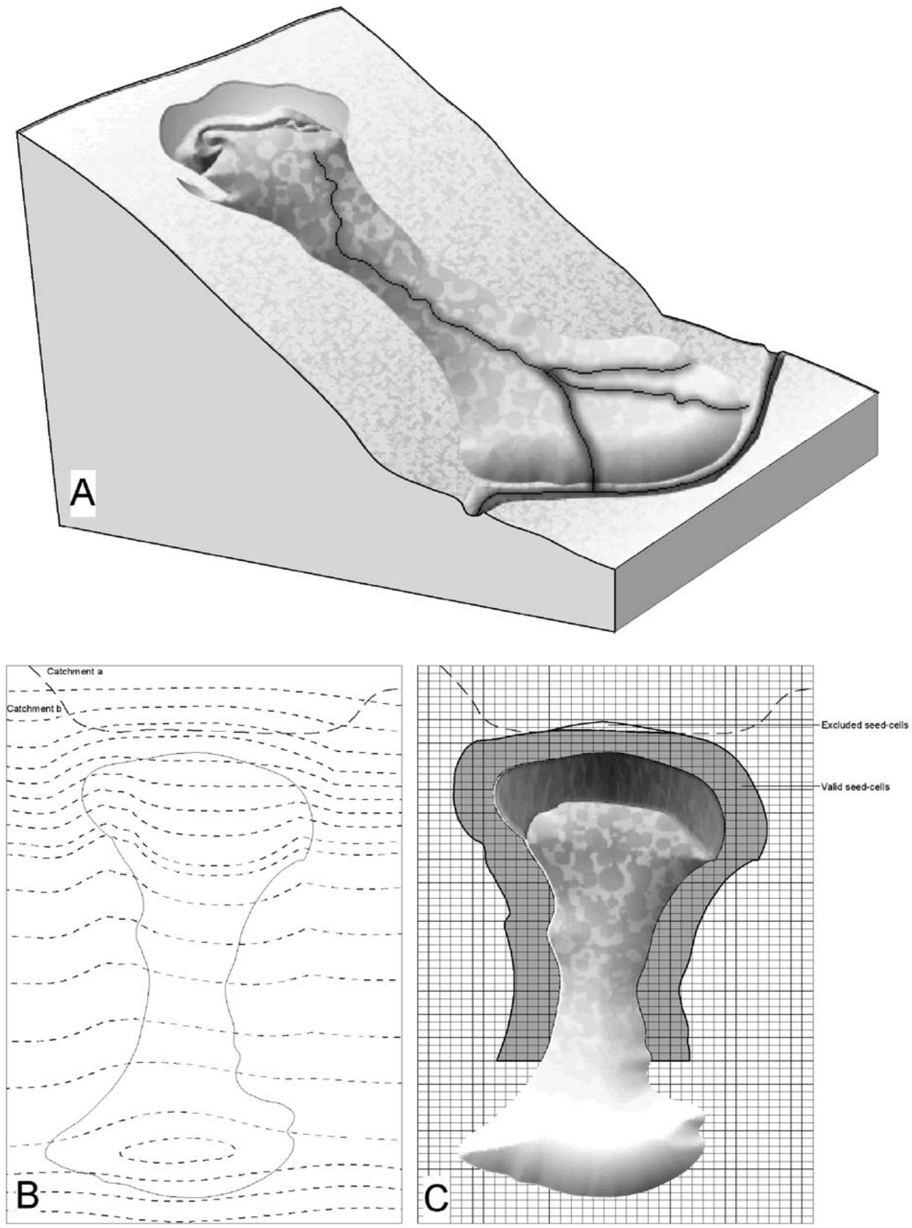



Fig. 5. Graphical representation of "Seed-cells" concept. (A) Block diagram of the investigated landslide. (B) Map representation of the same slide with contour lines and the catchment boundaries. (C) Seed-cell buffer zone with excluded cells passing over the other catchment and the valid seed cells.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

given as a sample of the 13 maps in Fig. 4. The produced maps and their thematic information considered in this study are presented in Table 1.

At the first glance, some parameter couples seem to provide redundant and duplicate information, such as drainage density–distance to drainage and distance to fault line–fault density couples. However, the elements of these couples represent different information about the area, such that the density functions explore through the data in kilometer squares which could be attributed to reflect the regional setting and, on the other hand, the distance functions that work in meter scales tend to evaluate the local situations. The density of fault map involves the probability of the presence of micro tremors and gives an indication of the degree of shearing of the rocks, whereas the distance to fault map indicates the presence of joints-fractures affecting the strength (Tansi et al., 2000) and increasing the probability of the presence of springs through the separated joints-fractures if not sealed. On the other hand, for drainage couple, the corrected distance to stream indicates the possibility of river bank erosion in the toe part of a landslide where the drainage density gives some clues about the regional hydrogeological properties of the rock units.

Upon completion of the parameter maps, the decision rules of landslide occurrences should have to be created. There have been numerous approaches in the literature to define the decision rules from feature-related spatial districts or ways of subdividing the terrain into equal conditions or by means of morphometrical units, such as unique condition areas, slope facets and geomorphological terrain units (Turrini and Visintainer, 1988; Gupta and Joshi, 1990; Carrara et al., 1991; Pachauri and Pant, 1992; Van Westen, 1993; Chung et al., 1995; Soeters and van Westen, 1996; Gupta and Anbalagan, 1997; Guzzetti et al., 1999). However, when considered in the spatial domain, these different methodologies involve both the responsible pixels and the irrelevant pixels about the landslide occurrence, which degrades the quality of the decision rules in the susceptibility assessment scheme as the presence of irrelevant pixels disseminate or contaminate the power of effective parameter ranges in the decision rules.

A new approach is followed in the generation of decision rules of landslide occurrence, called "seed cells", because it is considered that the best undisturbed morphological conditions (conditions before landslide occurs) would be extracted from the vicinity of the landslide polygon itself. This is achieved by adding a buffer zone to the crown and flanks of the landslide. The extraction rule for the boundaries of the buffer line is as follows: If the distance between the slide boundary and microcatchment divide line is smaller than 100 m, then use the microcatchment divide line; if the distance is larger than 100 m, use the 100-m buffer line (four seed cells) for the seed zone generation (Fig. 5). The distance of 100 m is taken as an arbitrary value but after the extraction of seed cells, a 150-m zone (six seed cells) is tested and it is seen that most of the cells fall beyond the microcatchment divides and hence are eliminated.

The decision rules should be created from the landslide-related pixels in the whole study area to reflect the real effect of parameter maps over the distribution of the phenomena. To store the landslide-related information into a database, a point mesh grid of 25 m is laid over the region. The seed-cell zones and landslide polygons are extracted from this grid and all of the information contained in 13 maps are transferred to the cells and stored respectively as "seed cell" and "slided mass" databases for further use of external statistical packages. The "seed cell" database will serve as a foundation for decision rule generation where the pre-slide conditions are simulated.

## 3. Statistical data preparation

Because the aim of this study was to increase the objectivity of the assessment, and to let the data derive its own decision rules, statistical methods are implemented. To achieve this goal, the nodes of seed cells are used as decision rule generators or training samples. Although there are 49 samples (landslides) in the area, a small population for statistical analyses, the conversion of this polygon data into seed cells results in a number of 4430 samples (seed cells) which is large enough to treat as a population. These 4430 samples contain the values of all of the 13 parameters, which are known to result in slope instabilities.

However, a major impediment is compromised in the nature of the data because all of the bivariate methods used in literature are designated to be used with some form of categorical data not to be used with continuous data sets. This arouses from the fact that the all-available bivariate methods base themselves onto the landslide density or abundance in certain parameter classes. If the continuous data are used, the densities will be calculated for the whole map and not even a single natural preference in the area will be utilized for susceptibility assessment. Consequently, a continuous to discrete map conversion seems to be indispensable. Some efforts have been carried out to categorize continuous data. Generally proposed meth-

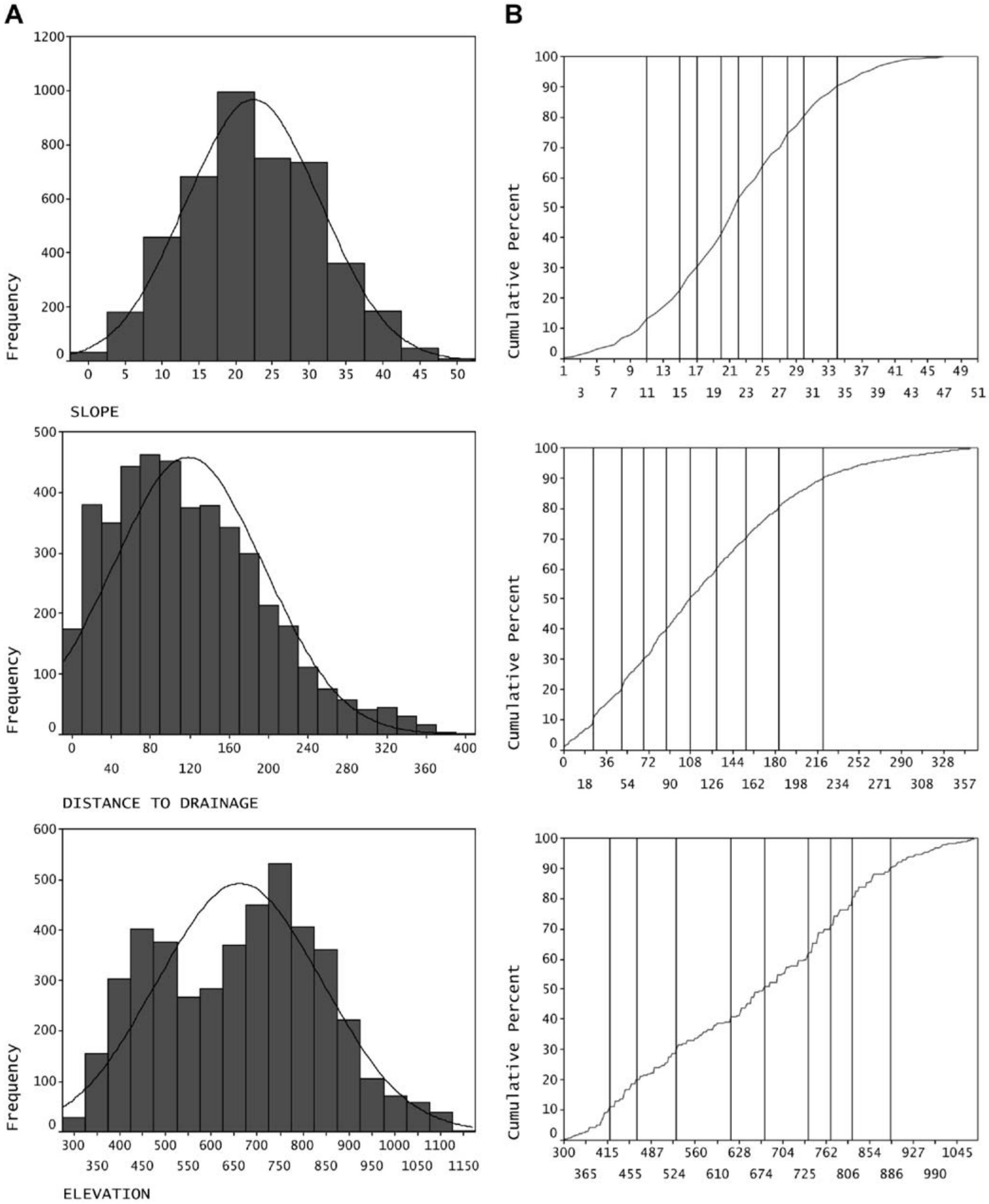



Fig. 6. Samples of percentile class limits on different distributions. (A) Original seed-cell distribution. (B) Calculated percentile classes shown on cumulative frequency diagrams of corresponding distribution.

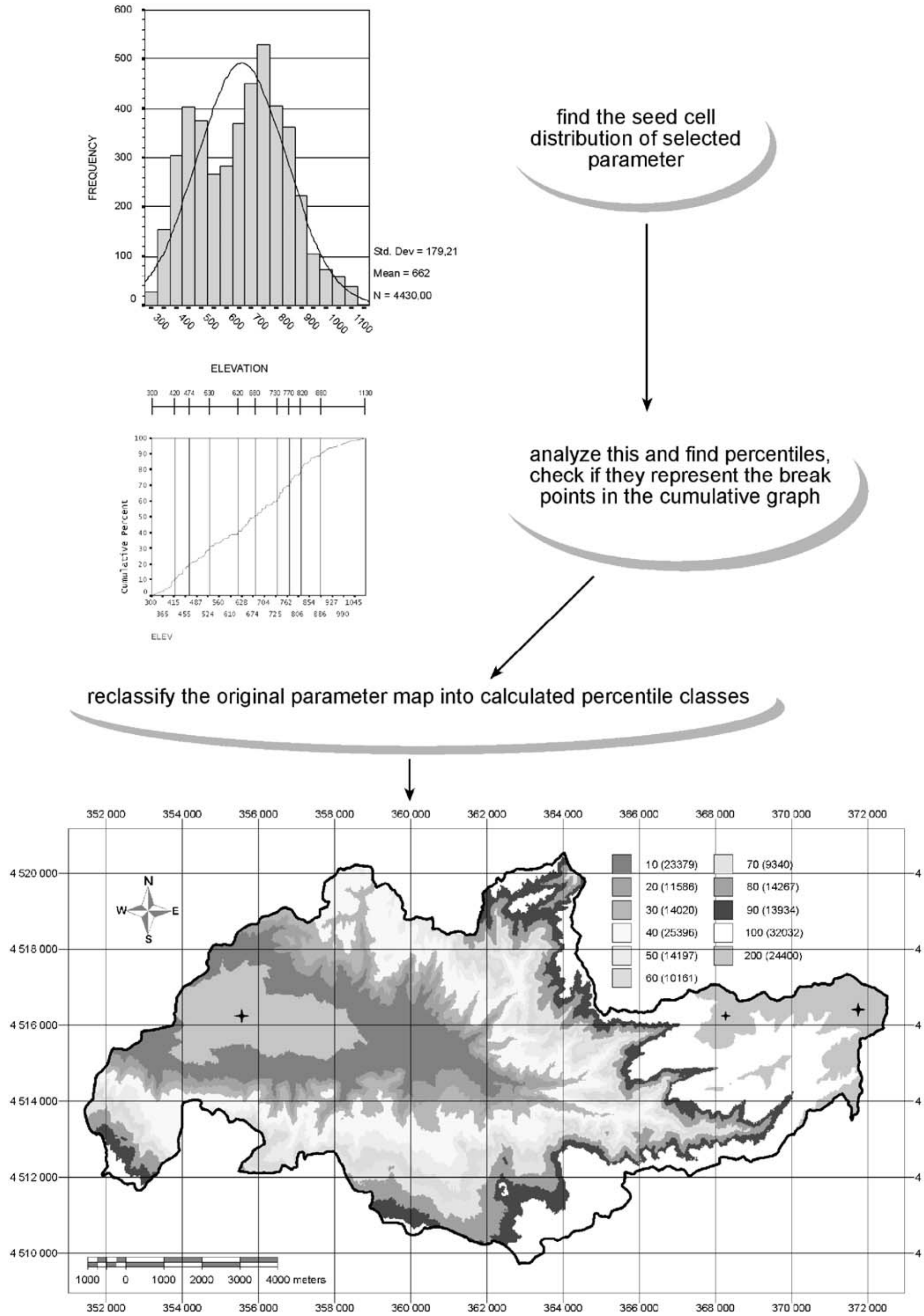



Fig. 7. The snapshot of methodology of percentile method and reclassified parameter map production. The areas with a star are the excluded areas because there are no landslides in these areas.

Table 2
The percentile values of seed cells within each variable

| | | ASPECT | SLOPE | ELEV | D_DRAIN | D_E5 | D_SETTLE | DENS_FAULT | D_FAY | D_RIDGE | D_PRO | DENSDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *N* | Valid | 4430 | 4430 | 4430 | 4430 | 4430 | 4430 | 4430 | 4430 | 4430 | 4430 | 4430 |
| | Missing | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mean | | 168.37 | 22.41 | 661.84 | 118.30 | 1217.83 | 699.45 | 235.56 | 408.52 | 113.88 | 176.22 | 160.68 |
| Median | | 180.00 | 22.00 | 680.00 | 108.00 | 1206.00 | 459.00 | 209.00 | 306.00 | 96.00 | 145.00 | 157.00 |
| Mode | | 45 | 22 | 750 | 25 | 1 | 153 | 0 | 114 | 13 | 9 | 150 |
| Standard deviation | | 126.25 | 9.12 | 179.21 | 76.99 | 899.59 | 651.25 | 161.03 | 339.66 | 82.12 | 142.29 | 37.01 |
| Variance | | 15,939.83 | 83.13 | 32,116.12 | 5927.73 | 809,255.17 | 424,130.59 | 25,930.03 | 115,368.76 | 6743.79 | 20,245.18 | 1369.43 |
| Range | | 359 | 50 | 830 | 397 | 3112 | 2625 | 705 | 1517 | 360 | 683 | 232 |
| Minimum | | − 1 | 1 | 300 | 0 | 0 | 4 | 0 | 0 | 4 | 0 | 55 |
| Maximum | | 358 | 51 | 1130 | 397 | 3112 | 2629 | 705 | 1517 | 364 | 683 | 287 |
| Percentiles | 10 | 9 | 11 | 420 | 25 | 75 | 105 | 12 | 73 | 14 | 17 | 119 |
| | 20 | 27 | 15 | 474 | 49 | 211 | 167 | 110 | 127 | 39 | 41 | 136 |
| | 30 | 45 | 17 | 530 | 68 | 391 | 230 | 157 | 180 | 52 | 70 | 144 |
| | 40 | 83 | 20 | 620 | 87 | 834 | 322 | 189 | 239 | 75 | 107 | 150 |
| | 50 | 180 | 22 | 680 | 108 | 1206 | 459 | 209 | 306 | 96 | 145 | 157 |
| | 60 | 225 | 25 | 730 | 130 | 1607 | 609 | 242 | 386 | 122 | 189 | 165 |
| | 70 | 284 | 28 | 770 | 155 | 1857 | 804 | 286 | 478 | 152 | 241 | 174 |
| | 80 | 315 | 30 | 820 | 183 | 2114 | 1397 | 357 | 675 | 191 | 302 | 187 |
| | 90 | 333 | 34 | 880 | 221 | 2427 | 1825 | 481 | 977 | 238 | 382 | 208 |

ods depend on the optimum bin width classification of the histograms. The earliest published rule for selecting the bin width appears to be that of Sturges (1926). This proposal is more of a number-of-bins rule rather than a bin width rule itself, but essentially the data range to choose the bin width.

$$\hat{h} = \frac{\text{Range of Data}}{1 + \log_2 n} \tag{1}$$

where $\hat{h}$ is the optimum bin width and $n$ is the sample size. However, Scott (1979, 1992) showed that this bin width leads to an oversmoothed histogram, especially for large samples, and proposed an unbiased estimation of a probability density function, which is achieved when:

$$W = 3.49\sigma N^{-1/3} \tag{2}$$

where $W$ is the width of the histogram bin, $\sigma$ is the standard deviation of the distribution and $N$ is the number of available samples. This estimator worked well for Gaussian distributions, where it led to overlay large bin widths and hence oversmoothing. Friedman and Diaconis (1981) suggested a more simple method:

$$W = 2(\text{IQR})N^{-1/3} \tag{3}$$

where $W$ is the width of the histogram bin, IQR is the interquartile range (the 75th percentile minus the 25th percentile) and $N$ is the sample size. Numerical comparisons by Emerson and Hoaglin (1983) of the Scott and Freedman–Diaconis (FD) rules showed that the FD rule led to narrower bin widths, although in practical applications, the two rules will often lead to the same choice of interval width (Izenman, 1991).

All of the methods cited above have a number of disadvantages, such as oversmoothed class divisions, distribution and sample size dependency, validity only for one parameter map and ineffective application to multi-modal distributions (i.e., multi-modality overrides the assumptions).

Also the same type of discussions remains always unclear in susceptibility analysis literature because most of the authors use their expert opinion for the boundaries of the classes (Carrara and Merenda, 1974; Malgot and Mahr, 1979; Rupke et al., 1988; Pachauri and Pant, 1992; 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 same data sets somewhere else.

A new data driven methodology is proposed in this study to overcome some of the disadvantages mentioned above. The continuous data sets are classified into categories based on percentile divisions of seed cells. Because the number of samples in all of the 13 maps is the same (4430) and because the parameters portray different distribution types, the effect of sample number and distribution type should be minimized in selecting the category limits. This operation is somewhat similar to histogram equalization in image processing, the existing values are however preserved. Thus, the starting point of this method is to define the group limits of equal frequencies, where frequency domain is quite free from distribution parameters and could easily be implemented in bimodal or multi-modal distributions. The examples of application of percentile division in different distributions and the resulting percentile limits are shown in Fig. 6.

The next step is to reclassify the original parameter maps into percentile maps by using simple logical query procedures inquiring the percentile limits. These steps are represented in Fig. 7, and carried out for the 11 continuous parameter maps. The remaining two categorical parameter maps (geological map and the land cover map) are excluded. Also the data range of the percentile map is dependent on the

Table 3
Methodological snapshot of landslide susceptibility

| |
|---|
| Bivariate analyses (landslide susceptibility) |
| $D_{\text{area}} = 1000\dfrac{\text{Npix}(SX_i)}{\text{Npix}(X_i)}$ |
| in which $D_{\text{area}}$ = areal density per mileage, Npix($SX_i$) = number of pixels with mass movements within variable class $X_i$, and Npix($X_i$) = 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: |
| $W_{\text{area}} = 1000\dfrac{\text{Npix}(SX_i)}{\text{Npix}(X_i)} - 1000\dfrac{\sum \text{Npix}(SX_i)}{\sum \text{Npix}(X_i)}$ |

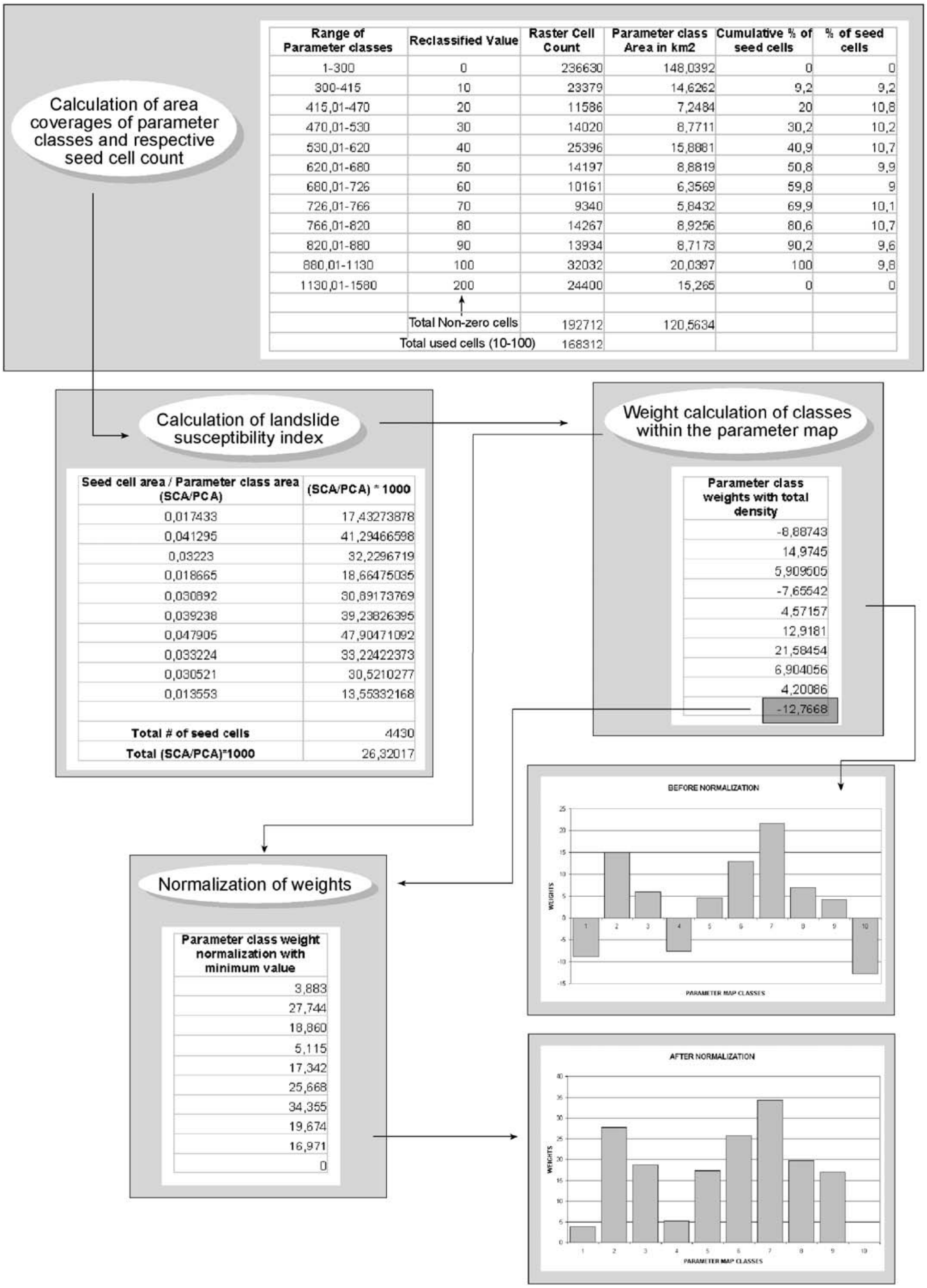



Fig. 8. The steps through landslide susceptibility analysis.

Table 4
Weight values of all available parameter classes

| | Parameter classes | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameter map | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
| Fault density | 0.26 | 0.00 | 9.65 | 23.87 | 37.26 | 32.81 | 26.93 | 41.21 | 33.73 | 34.72 |
| Elevation | 3.88 | 27.74 | 18.68 | 5.11 | 17.34 | 25.69 | 34.35 | 19.67 | 16.97 | 0.00 |
| Distance to ridge | 0.00 | 2.37 | 3.42 | 4.79 | 4.82 | 7.52 | 7.67 | 9.14 | 11.06 | 22.16 |
| Distance to settlement | 0.00 | 19.38 | 27.19 | 19.18 | 14.61 | 21.92 | 23.68 | 4.20 | 18.66 | 6.30 |
| Distance to power lines and road network | 5.53 | 3.65 | 2.10 | 2.54 | 6.90 | 9.19 | 11.70 | 14.30 | 15.55 | 0.00 |
| Distance to fault line | 0.00 | 9.67 | 15.76 | 17.16 | 17.02 | 16.11 | 17.62 | 3.89 | 4.79 | 4.76 |
| Distance to E-5 highway | 99.56 | 46.53 | 31.07 | 2.82 | 8.25 | 7.11 | 22.66 | 22.09 | 15.82 | – 0.01 |
| Distance to drainage | 0.00 | 0.35 | 4.44 | 7.67 | 8.72 | 9.50 | 11.46 | 16.00 | 19.30 | 19.83 |
| Drainage density | – 0.01 | 5.89 | 20.14 | 26.17 | 22.53 | 22.06 | 18.18 | 12.04 | 10.15 | 1.59 |
| Slope | 0.00 | 8.28 | 18.15 | 24.11 | 29.40 | 29.92 | 30.04 | 32.92 | 28.57 | 30.51 |
| Aspect | 23.33 | 32.02 | 26.71 | 18.51 | – 0.01 | 3.68 | 3.52 | 11.20 | 15.72 | 22.84 |
| | quat | BGY | DSA | tkbf | talus | bolugran | Eocay | sok | | |
| Geology | 0.67 | 2.08 | 58.04 | 35.95 | 3.69 | 28.73 | 22.68 | 19.60 | | |
| | Dense forest | Forest | Mixed | Settlement | Road | | | | | |
| Land cover | 4.69 | 29.03 | 14.43 | 0.16 | 124.04 | | | | | |

seed cells database, the data outside the seed cell buffer range are discarded because there are no landslides occurring in these regions.

To reduce complexity in the bivariate analysis stage, arbitrarily 10 classes or, in other words, 10 percentiles resulting in 10% divisions of the seed cells are arbitrarily proposed to use, because even in 10 classes with 11 parameters (excluding the 2 categorical ones), 110 different classes (11 maps × 10 class) should have to be maintained. However, it is obvious that lower class numbers would result in large generalizations in the final susceptibility map, where higher class numbers would result in isolated pixels that should have to be smoothed out; this would alter the result of the susceptibility map.

Table 2 represents the calculated percentile values of seed cells that are used in bivariate analyses.

### *3.1. Bivariate (landslide susceptibility) analysis and map creation*

In bivariate analyses, the core of the analysis is to measure the abundances of landslide occurrences within each percentile class. This is done by determining the total number of seed cells within each percentile class, then dividing this value by the total number of pixels within that percentile. This ratio gives the representative density of that percentile class because both the area represented by a seed cell and by a pixel is equal (25 × 25 m).

The equal frequency or percentile divisions of a parameter yields in same feature densities in each

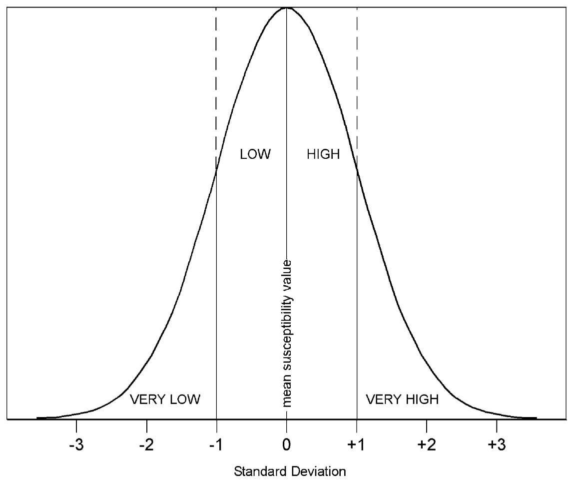
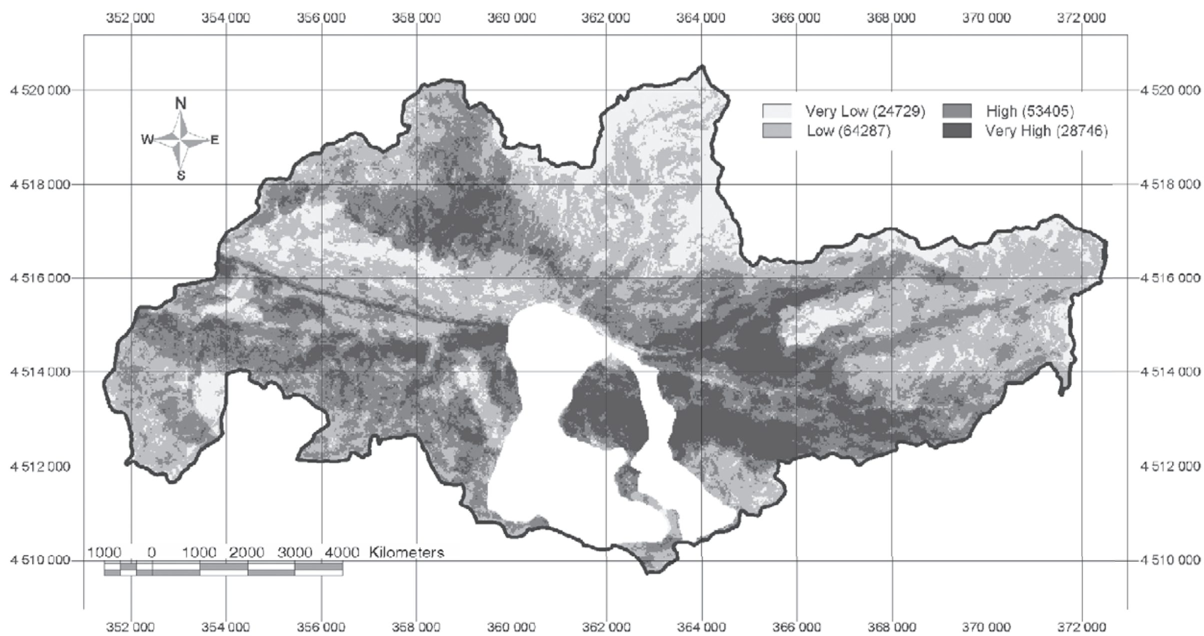



Fig. 9. Limits of susceptibility hazard classes on normal distribution.

class. As previously noted, the landslide abundances in each percentile class are set by the percentile method as approximately to 10% in each class; hence, the resultant percentile classes of the parameter map have the same number of seed-cell occurrence; however, the areas of these percentile classes on the map are not equal to each other, that is, densities are different. This constitutes somewhat a natural weighting of each responsible parameter class and parameter itself. In other words, this data-dependent categorization reduces the problem, in bivariate analyses, of what weight should be given to each parameter map, because each percentile class acts now like an individual parameter map and the ratio of the seed-cell abundance over the class area gives its natural weight.

A further step is to distribute the weights onto the whole map regarding all of the percentile classes of a parameter. In other words, to evaluate the influence of each parameter, weighting factors should be calculated to compare the calculated density of a percentile class with the overall density in the parameter map of the whole catchment area (Van Westen, 1993).

This is achieved by adding up all of the calculated weights of the percentile classes of a parameter map and then subtracting it from the individual percentile class's weight (Table 3). If the result of this subtraction yields a negative number, the effect of that percentile class can be considered not to have any influence in the stability of the slope; on the other hand, the positive values indicate a definite contribution of that particular percentile class to the landslide




| DN Value | Hazard Class | % area covered | % seed cells |
|---|---|---|---|
| 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. 10. The hazard map and the amounts of landslides in each class as a result of bivariate analysis.

occurrences. This process enables the percentile class weights to be comparable with each other. Here forth, any randomly selected percentile class of any map has a weight comparable with that of any other percentile class of any other map. The mathematical syntax of the method used is given in Table 3.

In the case of the presence of negative class weights, a minor normalization or rescaling step is required. This is done by adding the absolute values of greatest negative weight to all of the percentile classes for each particular parameter map. The steps of weight calculation and normalization processes are presented in Fig. 8. These steps are carried out for all parameter maps of the Asarsuyu catchment. The calculated weight values of all parameter classes are shown in Table 4, which are then used to construct the susceptibility map.

After the calculation of the weight values, for each percentile class within each parameter map, they are spatially summed up to yield the final pixel scores for the susceptibility map. No extra weighting procedure is applied in the summation process because the classes have already been normalized.

The resultant pixel values of the susceptibility map are then reclassified into any number of susceptibility classes. In this study, four classes were selected as very low, low, high, very high using the distribution parameters of resultant susceptibility map and the mean value of the susceptibility values is taken as the pivot point, then + and − standard deviations of the distribution are taken as boundaries of the classes, which is valid for normal or normal-like distributions (Fig. 9).

The resultant map and the landslide amounts in these susceptibility classes are given in Fig. 10. It is seen that 48% of the total area is classified as high and very high susceptibility class and within these classes, 93.3% of seed cells are encountered. On the other hand 52% of the study area is classified as low and very low susceptibility, which in turn hosts only 6.7% of total seed cells in the area.

## 4. Discussions and conclusions

Throughout this study, new approaches for existing bivariate statistical methods in landslide susceptibility assessment are proposed by increasing the data dependency of the models while decreasing expert knowledge demand. In addition, in the conversion of non-attributed vector data into continuous variables, true distance approach is implemented rather than map distances. This gradient correction is needed to portray the real conditions near a mapped feature, acknowledged with its limitations. The main limitation is that the variables produced from such a correction can only show local conditions, so they should not have to be considered as regional variables. The effect of a certain spatial feature to its neighborhood will certainly decrease with increasing distance and become negligible beyond a certain point. On the other hand, to assess the regional effects of the same feature, its density is implemented instead of its true distance.

Furthermore, in the generation of decision rules of landslide occurrence, a buffer zone is added to the crown and flanks of the landslides, to reveal the best undisturbed morphological conditions, simulating presliding conditions, to be stored as attributes of a new database called ‘‘Seed-Cell Database’’ which is used to extract the critical parameter ranges responsible for the landsliding mechanism. Conceptually, this seed-cell implementation has a limitation concerning the slope profile and the instability type. Rectilinear slopes and shallow flows or translational instabilities are the most suitable situations to apply this new idea because the slope form does not change drastically after a slide. Implementations for other types of slides and slope profiles should have to be carried out considering the slide type and resultant topographical changes.

In addition, while categorizing classes of parameter maps, a new approach is also engaged and presented as percentile method, which yields in data-dependent parameter class weightings in bivariate landslide susceptibility assessment procedures. The core of this methodology is composed of equal frequency class divisions of each parameter, rather than expert knowledge.

The produced susceptibility map is then tested in the field by the slope instabilities that occurred after the 12 November 1999, Kaynasli Earthquake. Approximately 20 shallow earth flows and translational landslides are observed as either new or reactivated slides in the field and all of these slides fall into high and very high classes of the produced susceptibility

Fig. 11. Samples of landslides in the study area after 12 November 1999, Kaynasli Earthquake.

map (Fig. 11). This validates the applicability of proposed methods, approaches and classification scheme.

## Acknowledgements

This study is supported by the Middle East Technical University Research Project No AFP-99-03-09-04. The authors thank Dr. Tamer Topal for his critical suggestions about the manuscript. Also, the great work of Dr. Nuretdin Kaymakçı in the modification of the manuscript is appreciated. Dr. Sorriso Valvo and the other anonymous referee are greatly appreciated for contributions in the improvement of the manuscript.

## References

Aktimur, T., Algan, Ü., Ateş, Ş., Oral, A., Ünsal, Y., Karatosun, H., Öztürk, V., Sönmez, M., 1983. Bolu ve yakın çevresinin yerbilimleri sorunları ve muhtemel çözümleri. MTA Rap No. 7387 (unpublished).

Aydın, M., Serdar, S., Şahintürk, Ö., Yazman, M., Çokuğraş, R., Demir, O., Özçelik, Y., 1987. Çamdağ (Sakarya)—Sünnicedağ (Bolu) yöresinin jeolojisi. TJK bülteni 30, 1–14.

Batum, İ., 1968. 1/25,000 ölçekli Adapazarı G26-b3 paftasının Jeolojisi. MTA Rap No. 4778 (unpublished).

Blumenthal, M., 1948. Bolu civarı ile Aşağı Kızılırmak mecrası arasındaki Kuzey Anadolu silsilelerinin jeolojisi. MTA Yayınları, Seri B, No. 13. 265 pp.

Bonham-Carter, G.F., 1996. Geographic Information Systems for Geoscientists, Modelling with GIS. Pergamon Press, Canada. 398 pp.

Brabb, E.E., 1984. Innovative approaches to landslide hazard and risk mapping. Proc., Fourth International Symposium on Landslides, vol. 1. Canadian Geotechnical Society, Toronto, Canada, pp. 307–324.

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., 1991. GIS techniques and statistical models in evaluating landslide hazard. Earth Surface Processes and Landforms 16 (5), 427–445.

Cerit, O., 1990. Bolu Masifinin jeolojik ve tektonik incelemesi. Doktora Tezi, Hacettepe Ü. 217 pp.

Chung, C.F., Fabbri, A.G., van Westen, C.J., 1995. Multivariate regression analyses for landslide hazard zonation. In: Carrara, A., Guzzetti, F. (Eds.), Geographical Information Systems in Assessing Natural Hazards. Kluwer Academic Publishers, Netherlands, pp. 107–134.

Dalgıç, S., 1994. Anadolu otoyolu Bolu dağı geçişinin mühendislik jeolojisi. Doktora Tezi, İstanbul Üniversitesi, Fen bilimleri Enstitüsü. 213 pp. (unpublished).

Emerson, J.D., Hoaglin, D.C., 1983. Stem and leaf displays. In: Hoaglin, D.C., Mossteiler, F., Tukey, J.W. (Eds.), Understanding Robust and Exploratory Data Analysis. Wiley, New York.

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

Fiener, Y.A.O., Haji, A.F., 1999. Application of hazard and risk maps for highway slopes management and maintenance. In: Yagi, N., Yamagami, T., Jiang, J.C. (Eds.), Slope Stability Engineering, vol. 2. Balkema, Rotterdam, pp. 1287–1290.

Friedman, D., Diaconis, P., 1981. On the histograms a density estimator $L_2$ theory. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 57, 453–476.

Gedik, A., Alkaş, İ., 1996. Bolu yöresinin Jeolojisi ve karbondioksid olanakları. MTA rap No. 10147 (unpublished).

Gupta, P., Anbalagan, R., 1997. Slope stability of Tehri dam reservoir area, India using landslide hazard zonation (LHZ) mapping. Quarterly Journal of Engineering Geology 30, 27–36.

Gupta, R.P., Joshi, B.C., 1990. Landslide hazard zoning using the GIS approach—a case study from the Ramganga catchment, Himalayas. Engineering Geology 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.

Işın, Z.S., 1999. D-100 (E-5) Karayolu, Bolu Dağı geçişi, Kaynaşlı—Elmalık arasının mühendislik jeolojisi ve Taşaltı Heyelan Bölgesinin jeoteknik değerlendirilmesi. Yüksek Lisans Tezi, İ.T.Ü., 145s.

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

Izenman, A.J., 1991. Recent developments in non-parametric density estimation. Journal of the American Statistical Association 86 (413), 205–224.

Ketin, İ., 1955. Akçakoca-Düzce bölgesinin jeolojik lövesi hakkında memuar. MTA Rap No. 2277 (unpublished).

Kienholz, H., 1978. Maps of geomorphology and natural hazards of Grindelwald, Switzerland, scale 1:10.000. Arctic and Alpine Research 10, 169–184.

Kobashi, S., Suzuki, M., 1988. Hazard index for the judgement of slope stability in the Rokko Mountain region. Proc., Interpraevent, 1988, Graz, Austria, vol. 1, pp. 223–233.

Malgot, J., Mahr, T., 1979. Engineering geological mapping of the west Carpathian landslide areas. Bulletin of the International Association of Engineering Geology 19, 116–121.

Orkan, N.İ., Aktimur, H., Sungur, G., Işıklar, İ.S., 1977. Ankara–İstanbul otoyolu güzergahı Bolu dağı jeoloji incelemesi. MTA Rap No. 2277 (unpublished).

Pachauri, A.K., Pant, M., 1992. Landslide hazard mapping based on geological attributes. Engineering Geology 32, 81–100.

Rupke, J., Cammeraat, E., Seijmonsbergen, A.C., van Westen, C.J., 1988. Engineering geomorphology of the Widentobel catchment, Apenzell and Sankt Gallen, Switzerland. A geomorphological inventory system applied to geotechnical appraisal of slope stability. Engineering Geology 26, 33–68.

Scott, D.W., 1979. On optimal and data-based histograms. Biometrika 66 (3), 605–610.

Scott, D.W., 1992. Multivariate Density Estimation. Wiley, New York.

Soeters, R., van Westen, C.J., 1996. Slope instability recognition analysis and zonation. In: Turner, K.T., Schuster, R.L. (Eds.), Landslides: Investigation and Mitigation. Transportation Research Board National Research Council, Special Report, vol. 247, pp. 129–177. Washington, DC.

Sözen, A., Erbayar, M., Çamaşırcıoğlu, A., Çeltek, N., 1996. Düzce (Bolu)—Devrek (Zonguldak) yöresinin genel jeokimya raporu. MTA Rap No. 9922 (unpublished).

Sturges, H.A., 1926. The choice of a class interval. Journal of the American Statistical Association 21, 65–66.

Süzen, M.L., 2002. Data driven landslide hazard assessment using geographical information systems and remote sensing, M.E.T.U. PhD Thesis, Unpublished. 196 pp.

Süzen, M.L., Yeşilnacar, E., 2002. A landcover extraction scheme for landslide hazard assessment using satellite image components. International Conference on Environmental Problems of the Mediterranean Region at Near East University, Nicosia, North Cyprus, 11–16 April. In Press.

Tansi, C., Sorriso-Valvo, M., Greco, R., 2000. Relationships between joint separation and faulting: an initial numerical appraisal. Engineering Geology 52, 225–230.

Turrini, M.C., Visintainer, P., 1988. Proposal of a method to define areas of landslide hazard and an application to an area of the dolomites, Italy. Engineering Geology 50, 255–265.

Uromeihy, A., Mahdavifar, M.R., 2000. Landslide hazard zonation of the Khorshostam area, Iran. Bulletin of Engineering Geology and the Environment 58, 207–213.

Van Westen, C.J., 1993. Application of geographic information systems to landslide hazard zonation. ITC Publication, vol. 15. International Institute for Aerospace and Earth Resources Survey, Enschede, The Netherlands. 245 pp.

Yin, K.L., Yan, T.Z., 1988. Statistical prediction model for slope instability of metamorphosed rocks. In: Bonnard, C. (Ed.), Proc., Fifth International Symposium in Landslides, Lausanne, vol. 2. A.A. Balkema, Rotterdam, Netherlands, pp. 1269–1272.