Earth Surface Processes and Landforms
Earth Surf. Process. Landforms **26**, 1251–1263 (2001)
DOI: 10.1002/esp.263

ASSESSMENT OF SHALLOW LANDSLIDE SUSCEPTIBILITY BY MEANS OF MULTIVARIATE STATISTICAL TECHNIQUES

CRISTINA BAEZA* AND JORDI COROMINAS
Department of Geotechnical Engineering and Geosciences, Technical University of Catalonia (UPC), Jordi Girona 1-3, Edifici D-2, 08034 Barcelona, Spain

Received 13 March 2000; Revised 8 April 2001; Accepted 7 June 2001

ABSTRACT

Several multivariate statistical analyses have been performed to identify the most influential geological and geomorphological parameters on shallow landsliding and to quantify their relative contribution. A data set was first prepared including more than 30 attributes of 230 failed and unfailed slopes. The performance of principal component analysis, t-test and one-way test, allowed a preliminary selection of the most significant variables, which were used as input variables for the discriminant analysis. The function obtained has classified successfully 88·5 per cent of the overall slope population and 95·6 per cent of the failed slopes. Slope gradient, watershed area and land-use appeared as the most powerful discriminant factors. A landslide susceptibility map, based on the scores of the discriminant function, has been prepared for Ensija range in the Eastern Pyrenees. An index of relative landslide density shows that the results of the map are consistent. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS: landslide susceptibility; multivariate techniques; discriminant analysis

INTRODUCTION

The use of landslide susceptibility and hazard maps for land-use planning has increased significantly during the last few decades. The purpose of these maps is the identification of areas threatened by present and potential slope instability. Their reliability depends mostly on the amount and quality of available data used as well as on the selection of the appropriate methodology for susceptibility and hazard assessment. On the other hand, the working scale also affects the quality of the results (Van Westen, 1994). Comprehensive syntheses of methods for susceptibility and hazard assessment can be found in Hansen (1984), Hartlén and Viberg (1988) and Corominas (1992). Procedures can be grouped as based on geomorphological analysis, data treatment techniques and deterministic approaches. The first two are commonly used in regional hazard analyses, while the last one is used in detailed studies, in which the safety factor of the slopes is determined. The latter is probably the best hazard assessment approach as the physical principles of the slope stability are taken into account. However, the large amount of data that it requires can only be afforded in the case of individual slopes or small areas so it is not suitable when analysing large areas.

On a medium to small regional scale, landslide susceptibility is assessed by methods ranging from simple geomorphological analysis to complex data treatment. The reliability of the geomorphological analysis depends on the appropriate interpretation of the landscape, which is based on subjective expert criteria. In order to reduce subjectivity and quantify the degree of susceptibility, data treatment techniques have been incorporated. With these techniques, landslide-susceptible areas are determined by correlating some of the main factors that contribute to the occurrence of slope failure, such as steep slopes or presence of weak lithological units, with the past distribution of landslides. Several methods exist for establishing the correlation. The most simple consists of overlapping maps of the instability factors with the landslide inventory map (Nilsen *et al.*, 1979).

* Correspondence to: C. Baeza, Department of Geotechnical and Geosciences, Technical University of Catalonia (UPC), Jordi Girona 1-3, D-2 Building, E-08034 Barcelona, Spain. E-mail: cristina.baeza@upc.es
Contract/grant sponsor: Spanish Scientific Council-CICYT.

More sophisticated approaches of overlapping incorporate weighting procedures of the instability factors (Yin and Yan, 1988; Bonham-Carter *et al.*, 1990; Chung and Fabbri, 1993; Chung and Leclerc, 1994).

Among the data treatment approaches, multivariate analysis is one of the most sophisticated techniques for landslide susceptibility assessment. In the multivariate analysis, slope failure is considered as the result of the interplay of several interrelated environmental factors that can vary in space and time. Multivariate analysis allows the estimation of the relative weight of each contributing factor by means of statistical procedures such as multiple regression or discriminant analysis. These procedures have already been applied in landslide susceptibility assessment, demonstrating that they are capable of successfully predicting slope failures using either categorical data (Jones *et al.*, 1961; Kawakami and Saito *et al.*, 1984; Yin and Yan, 1988; Mora and Vahrson 1994) or quantitative terrain parameters (Carrara, 1983a,b; Mulder, 1991). However, the results obtained so far with these approaches show that less than 80 per cent of the land-units can be properly classified according to their degree of instability.

All data treatment approaches are very sensitive to the type and quality of the factors chosen for the susceptibility analysis. Very often, factors used are chosen because they can be easily gathered with current data-capture techniques rather than because they are the most suitable for the susceptibility analysis. Information about geology and land use can be easily found in many regions while slope morphometric parameters can be derived without difficulty from satellite imagery and digital terrain models. However, some factors which play a significant role in the stability of the slope, such as colluvium thickness or groundwater conditions, are often missing. In our opinion, the results of the aforementioned approaches may be improved by adding to the data set information relative to factors that are better linked to the actual behaviour of the slope.

The aim of our research has been the identification of the terrain attributes related to the occurrence of shallow landslides and to quantify their relative contribution to the instability of the slope. The most significant factors have been included in a discriminant function in order to define landslide susceptibility classes. To this purpose we have used a set of data that have been collected in the field. The validity of the discriminant function for landslide susceptibility assessment has been tested in a pilot area in the Spanish Eastern Pyrenees.

THE STUDY AREA

The methodology described in this paper has been developed using data from shallow landslides and stable slopes of the Spanish Eastern Pyrenees. The main landslide-triggering factor in the region is rainfall. Short-lasting rains of high intensity trigger debris flows and shallow slides while long-lasting rains of moderate intensity reactivate mostly earthflows and deep-seated landslides (Corominas and Moya, 1999). Heavy rains tend to occur during the autumn season (October and November) when warm and moist air masses coming from the Mediterranean Sea are forced to flow upwards through the Pyrenean mountains. If a cold air front is located beyond the range, the collision between the air masses causes sudden condensation and subsequent heavy precipitation. Rainfall records of over 200 mm in 24 h are usual in such circumstances and precipitation of over 700 mm per day was recorded during an extreme event in 1940 (Novoa, 1984).

In order to carry out the statistical analysis a data set has been prepared by collecting terrain attributes in two areas of the Eastern Pyrenees (Figure 1). These areas were chosen because they were affected by intense landslide activity during an exceptional rainfall event in November 1982 (Gallart and Clotet, 1988; Corominas and Alonso, 1990). The magnitude of this event is reflected by the amount of rainfall recorded in 48 h, which reached for instance, 555·8 mm at La Molina and 340 mm at La Pobla de Lillet and triggered several thousands of shallow slides and debris flows throughout the region.

Geological and geomorphological characteristics of the area are described in previous works (Puigdefábregas *et al.*, 1979; Muñoz, 1985; Muñoz *et al.*, 1986). Most of the landslides developed on marly and clayey formations (Keuper, Upper Cretaceous, Lower Palaeocene and Lower Eocene) and especially on colluvium blanketed slopes.

METHODOLOGY

The assessment of the landslide susceptibility of the slopes has followed several steps. First of all, a database has been prepared, which includes more than 30 attributes of failed and unfailed slopes. The attributes were

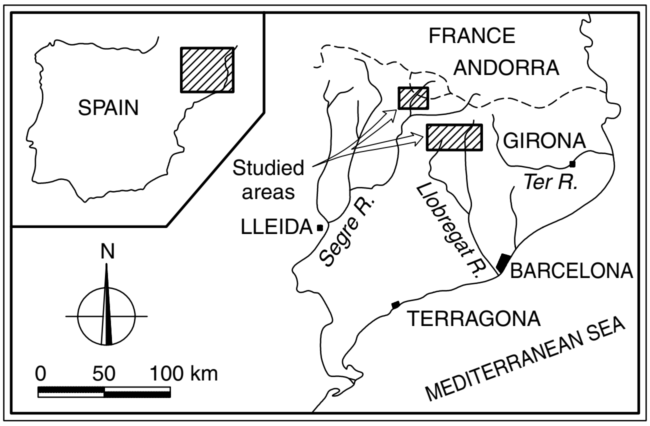

Figure 1. Location of the study area

chosen because of their expected relationship with the occurrence of the slope failures. These attributes have been used, directly or after being transformed, as variables for the statistical analysis. The latter has involved several tests for validating variables and a procedure for weighting and selecting the most significant variables, which will be finally included in the discriminant function. By this analysis, each considered slope takes the value given by a discriminant function. Ideally, failed and unfailed slopes should cluster around different values of the discriminant function, that is to say, the values of the function that each slope takes should indicate its degree of susceptibility to failure. The function (*DS*) is expressed by a combination of weighted variables:

$$DS = C_0 + C_1X_1 + C_2X_2 + \cdots + C_nX_n$$

where X is a variable contributing to instability, according to their statistical significance; C is a coefficient estimated in such a way that variability is maximal between failed and unfailed groups and minimal within each group. Thus, for a proper discrimination between failed and unfailed slopes, discriminant scores (*DS*) of each group should differ as much as possible. The standardized coefficients of C are used as the relative contribution of the variable to landslide susceptibility. The discriminant scores can be used to classify new slopes with unknown affiliation into one of the two groups. According to the discriminant score the slope is included within a landslide susceptibility class (Neuland, 1976; Carrara, 1983a). The classification technique is based on Bayes's rule, using a conditional probability. The percentage of slopes correctly classified is seen as an index of the effectiveness of the discriminant function. A detailed description of the discriminant analysis can be found in David *et al.* (1977), Lebart *et al.* (1982) and Dillon and Goldstein (1986).

The discriminant scores obtained by the slopes or land units in a given area, therefore, allow their rating in terms of their proneness to slope failure, thus enabling the establishment of susceptibility classes. The different steps are described in more detail below.

Data collection

The slope failures triggered by heavy rains in the Pyrenean region are not randomly distributed; they are frequent in hollows filled with colluvium and where the convergence of the groundwater flow is expected. Furthermore, failures appeared in steep slopes at some distance downhill from water divides. Therefore, a simple failure mechanism has been assumed from this description. The slope failure occurs in the susceptible sites due to the increase of pore-water pressures produced by rainfall infiltrated in the watershed with the subsequent reduction of the slope resisting forces. This mechanism has been proposed as the triggering factor for debris flows and shallow landslides (Johnson and Sitar, 1989). No water inflow coming from a neighbouring basin is assumed in our model.

An inventory including failed and unfailed slopes was first prepared. Failed slopes correspond to those triggered by heavy rains of November 1982 and were identified and located by means of aerial photographs taken in 1983. A total of 230 selected slopes were visited in the field in order to collect geological and morphological parameters that were considered relevant for the stability, which will be described later. The parameters were obtained directly in the field using a portable tape, a clinometer and an electronic distance meter. A standardized form was prepared for this purpose.

The failed slopes involve different types of landslides such as mudslides, shallow translational slides and debris flow of small size (volume of less than 10 000 m^3). The sample included a representative range of landslide types, lithology, land use and vegetation cover, that is to say, different environmental factors that can influence slope stability. The parameters collected were simple and easy to obtain in order to allow the inventory of a representative number of movements and slopes. Soil properties (grain size distribution, plasticity indexes and shear strength) were determined in the laboratory. However, the performance of more than 40 shear tests on undisturbed samples showed that peak and residual strength values corresponding to different lithostratigraphic units, overlapped, therefore, geotechnical parameters were not included as variables in the statistical analysis.

Variables used

Slope attributes collected encompass a variety of conditioning factors, namely geomorphology, lithology, hydrology, vegetation cover and morphometry of the slope. The slope attributes have been used, directly or transformed using simple mathematical operations, as variables for the multivariate analysis. The selected variables in this study are shown in Table I, and a brief explanation of their meaning follows.

Altitude (H): in several landslide events, it has been observed that the amount of precipitation and the number of landslides increases with altitude (Carrara, 1983a; Gallart and Clotet, 1988).

Aspect (α): the aspect of the slope has an indirect influence on moisture content of the soil, which is related to the reduction of the effective stresses at the potential failure surface (Neuland, 1976; Carrara, 1983a).

Slope angle (β): slope angle is probably the main factor of stability as it affects the magnitude of both normal and shear stresses on the potential surface of failure.

Watershed area (A): Zaruba and Mencl (1969) found a relationship between the occurrence of landslides and the size of watershed area. The larger the watershed area the larger the amount of water that infiltrates into the ground, thus increasing instability conditions (Okimura, 1983; Oyagi, 1984). Furthermore, the presence of colluvial deposits is more frequent in large watersheds.

Watershed angle (γ): gentle or low gradient watersheds allow greater rainfall infiltration and larger groundwater flow towards the failure zone (Corominas *et al.*, 1992; Baeza, 1994).

Length of watershed (L): has a similar role to watershed area (Oyagi, 1984; Corominas *et al.*, 1992).

Tree density (D): tree roots stabilize colluvial deposits thus restricting the occurrence of a slope failure. On the other hand, trees increase both water infiltration and evapotranspiration (Greenway, 1987; Mulder, 1991; Baeza, 1994). This parameter was obtained in the field by measuring the average distance between trees (dt) and was later transformed to tree density (number of trees in a 10 m × 10 m parcel) using the expression: $D = (10/dt)^2$.

Thickness of superficial deposits (Z): most shallow failures take place on superficial deposits, especially on colluvium. Thick deposits favour groundwater flow and the development of high pore-water pressures in the soil.

Cross-section of the slope (CRS): this variable shows the ability of topography to disperse or concentrate water to the slope, which can become more susceptible to landsliding (Smith, 1988; Gao, 1993; Baeza, 1994).

Slope complexity (SC): the slope complexity attempts to show how geologic structure may have an influence on the amount of water flowing towards the slope. It reflects the ability of the geological structure to feed the slope with water coming from neighbouring watersheds. For instance, the attitude of strata may influence the development of preferential groundwater paths (Baeza, 1994; Luzi, 1995).

Land use (LU): refers to the type of vegetation cover on the slope, which may stabilize colluvial deposits and affects the rainfall infiltration capacity of the slope. Sparsely vegetated areas show faster erosion and greater instability than dense forests (Anbalagan, 1992; Baeza, 1994).

Soil type (ST): lithology is one of the most influential parameters on slope stability, because each material has different shear strength and hydraulic conductivity. However, because we are dealing with shallow landslides, most of the inventoried failures involve colluvium and a few weathered bedrock formations.

Table I. Variables selected for statistical analysis

Variable	Value	Observations
Quantitative		
H, Altitude above mean sea level (m)		
α, Slope aspect (°)		
β, Slope angle (°)		
γ, Mean watershed (lg)		
L, Length of watershed (lg)		
A, Watershed area (lg)*		
D, Tree density (number of trunks in cells of $10\ \mathrm{m} \times 10\ \mathrm{m}$ in size)		
Z, Thickness of superficial deposits (lg)		
Qualitative		
Cross-section (*CRS*)		
convex	1	Lower values express higher capability to concentrate runoff and ground-water flow.
rectilinear	0	
concave	−1	
Slope complexity (*SC*)		
opposite	1	Higher values express higher capability of groundwater flow in reaching specific site on the slope.
two slopes	1·5	
flat /horizontal	4	
flat and	6	
same as topographic surface	10	
Land use (*LU*)		
bedrock	1	Higher values express higher infiltration and drainage capacity.
grassland	6	
shrubs	9	
clear forest	12	
dense forest	15	
scree deposits	21	
Soil type (*ST*)		
colluvium	2	Higher values express higher resistance to fail.
sandy	2·5	
clayey	3	
flysch	4·5	
bedrock	6·5	
Grouping variable (*SL*)		
failed slope	1	Variable only used in discriminant analysis to classify the sample.
unfailed slope	0	

* lg, logarithm

Type of slope (*SL*): failed and unfailed slope is used as a 'grouping variable' in discriminant analysis. Each slope was classified according to whether it failed or not during the 1982 rainfall event.

Statistical treatment

The statistical treatment was carried out using the SPSS Inc. (1988) statistical package. Successive steps were followed: (i) transformation of variables; (ii) testing for normal distribution; (iii) selection of independent variables; (iv) obtaining the discriminant function; and (v) testing of the discriminant function.

Transformation of variables

Some statistical techniques (i.e. factorial or discriminant) have difficulties in dealing with qualitative data. Even though qualitative variables can be included in discriminant analysis, optimal results are not guaranteed (Dillon and Goldstein, 1986). Therefore, using our expert criterion, we have given numerical values to qualitative variables (Table I).

Testing for normal distribution

Several statistical techniques require that variables be normally distributed. Each selected variable was tested for normal distribution using the Kolmogorov–Smirnov (K–S) test at 5 per cent confidence level. K-S compares the cumulative distribution function for a variable with a specific theoretical distribution using mean and standard deviation parameters. When the distribution of the sample is not symmetrical but positively skewed, this distribution can be transformed with logarithms to obtain a normal distribution. In our case, watershed area (A), thickness of superficial deposits (Z), and tree density (D) were transformed. Altitude (H) and slope aspect (α) were too biased to be successfully transformed and were finally rejected. Nevertheless those variables, which, because of them being discrete, usually do not present a normal distribution, were not initially rejected. Nie *et al.* (1981) proved that those considered to be relevant for the analysis could be analysed with no transformation using discriminant techniques.

Selection of independent variables

Performance of a linear discriminant function is likely to be poor when dealing with populations characterized by strong correlation between variables (Dillon and Goldstein, 1986). Consequently, dependent variables must be removed. Search for possible correlation has been performed with principal component analysis (PCA). This technique also provides some insight on the structure of the overall population. Knowing how the population is structured may allow the identification of groups of either variables or slopes having similar behaviour, to recognize some trends within the sample and to detect correlation that is difficult to observe with a simple correlation matrix. Besides PCA, two additional tests, t-test and one-way test, based on analysis of mean and variance, have been used. They provide an early understanding about the influence of each variable on stability. With the information obtained at this step it is possible to undertake a first selection of variables. Only independent variables showing the highest significance in relation to the slope stability were selected for discriminant analysis.

Table II presents the rotated matrix with seven factors that explain the 89·4 per cent of the variance. Weightings of the factors show that each factor is defined by only one variable, except for the first two factors. Variables L and A which express the capability to concentrate groundwater flow, define the first factor (F1); tree density (D) and land use (LU) define the second factor (F2); while lithology (Z and ST) and slope morphometry (SC, γ, β and CRS) define the remaining factors. There exists a distinct correlation between two pairs of variables in F1 (L and A) and F2 (D and LU), while the remaining variables are considered independents. In this case, we have rejected the correlated variable showing the lowest significance on the stability of the slope. t-Test and one-way test were performed to look for differences in mean and multiple variances for each variable in relation to the grouping variable (SL), which is defined as a binary variable (0 = unfailed slope; 1 = failed slope). Results displayed by t-test (Table III) show that SC (slope complexity) and ST (soil type) have low statistical significance, as the means of the unfailed slope and failed slope groups are very close. One-way test indicates which of the correlated variables best characterizes the stability of the slope. According to the F value we kept variable A ($F = 30·179$) in front of L ($F = 15·950$).

Table II. Rotated factor matrix with weights over 0·4 in bold

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
β	0·012	−0·117	0·033	−0·050	0·151	**0·977**	−0·025
γ	0·165	−0·011	0·085	−0·103	**0·942**	0·161	−0·021
L	**0·917**	0·014	0·064	−0·082	0·065	0·082	−0·064
A	**0·886**	0·130	−0·093	0·119	0·108	−0·064	−0·091
Z	**0·465**	0·201	**−0·563**	0·258	0·236	−0·027	−0·050
D	0·108	**0·914**	−0·016	−0·002	−0·051	−0·063	−0·023
CRS	−0·127	−0·010	0·026	−0·030	−0·020	−0·024	**0·990**
SC	0·038	0·007	−0·044	**0·975**	−0·096	−0·048	−0·029
LU	0·046	**0·920**	−0·005	0·023	0·047	−0·066	0·009
ST	0·057	0·049	**−0·923**	0·021	0·153	0·025	0·010

SC and *ST* with respectively $F = 0{\cdot}288$ and $F = 2{\cdot}085$, for a probability of 0·01 were rejected as well. F values of the other pair of correlated variables (*D*, *LU*) were so close ($F = 6{\cdot}696$ and $F = 7{\cdot}557$ respectively) that we decided to follow the procedure of discriminant analysis using both variables separately.

The significance of lithology is surprisingly low, given that it is usually one of the strongest factors affecting slope stability. This may be explained by the fact that shallow landslides affect mostly colluvial deposits rather than bedrock formations, and there is no significant variability of the strength properties of colluvium in the studied area. The homogeneity of the results of the shear tests, which has already been mentioned, also supports the lack of discriminant capability of the lithology.

Obtaining the discriminant function

From PCA, a data set including the most significant independent variables was taken as input of the discriminant analysis. In order to keep the independent variables assumption, the analysis was performed twice. The first one was made incorporating variable *D* and the second one, variable *LU*.

Assuming that the independent variables must have a multivariate normal distribution with equal variance–covariance matrix for each of the unfailed and failed slope groups, the first discriminant analysis was performed with the following input variables: β, γ, *A*, *Z*, *LU* and *CRS*.

The method of selecting variables for the discriminant analysis was the stepwise procedure, whereby variables are entered and removed one at a time from the discriminant function until the most significant model had been generated. The stepwise method selects variables by minimizing the sum of unexplained variation (residual variance) between groups previously identified by grouping variable (failed–unfailed). The discriminant function (DF1) obtained and the main statistical parameters are shown in Table IV. In this table the selection rule and statistical controls are summarized.

The magnitudes of the standardized discriminant weights indicate the contribution of each variable to the function. Slope gradient (β) is the most influential variable with a weight of 0·865, followed by watershed area (*A*) with 0·638 and land-use (*LU*) with −0·414. Transverse section of failure zone (*CRS*), mean watershed gradient (γ) and thickness of superficial deposits (*Z*), with lower weights, were also chosen.

Table III. T-test and one-way test results to compare between unfailed slope and failed slope groups

Variable	Slope*	Mean	Standard deviation	Standard error	F test ($\alpha = 0{\cdot}05$)	T test ($\alpha = 0{\cdot}01$)	F ratio	F prob. ($\alpha = 0{\cdot}01$)
β	0	26·151	6·735	0·952	$\sigma_0^2 = \sigma_1^2$	$\mu_0 \neq \mu_1$	54·749	0·000
	1	34·431	6·964	0·549				
γ	0	25·371	7·938	1·123	$\sigma_0^2 \neq \sigma_1^2$	$\mu_0 \neq \mu_1$	4·865	0·028
	1	28·857	10·256	0·808				
L	0	1·673	0·378	0·053	$\sigma_0^2 \neq \sigma_1^2$	$\mu_0 \neq \mu_1$	15·950	0·000
	1	2·020	0·577	0·046				
A	0	3·020	0·445	0·063	$\sigma_0^2 \neq \sigma_1^2$	$\mu_0 \neq \mu_1$	30·179	0·000
	1	3·529	0·607	0·048				
D	0	0·219	0·150	0·021	$\sigma_0^2 = \sigma_1^2$	$\mu_0 \neq \mu_1$	14·444	0·000
	1	0·317	0·161	0·013				
Z	0	0·375	1·207	0·171	$\sigma_0^2 \approx \sigma_1^2$	$\mu_0 \neq \mu_1$	6·696	0·010
	1	−0·059	0·977	0·077				
CRS	0	0·180	0·850	0·120	$\sigma_0^2 = \sigma_1^2$	$\mu_0 \neq \mu_1$	8·954	0·003
	1	−0·201	0·744	0·063				
SC	0	2·810	1·714	0·242	$\sigma_0^2 = \sigma_1^2$	$\mu_0 = \mu_1$	0·288	0·592
	1	2·978	1·999	0·158				
LU	0	11·700	4·122	0·583	$\sigma_0^2 = \sigma_1^2$	$\mu_0 \neq \mu_1$	7·557	0·006
	1	9·894	4·037	0·318				
ST	0	3·200	1·064	0·151	$\sigma_0^2 = \sigma_1^2$	$\mu_0 = \mu_1$	2·085	0·150
	1	2·956	1·034	0·082				

* 0, unfailed; 1, failed.

Table IV. DF1 and DF2 discriminant function results

Variables	Function coefficients: Standard	Unstandard	Correctly classified (%) (DI = −0·4)
DF1*			
β	0·865	0·124	General, 88.5
A	0·638	1·170	Partial (unfailed–failed), 81·4–95·6
LU	−0·414	−0·102	
TS	−0·237	−0·298	
γ	−0·213	−0·023	Group centroids
Z	0·204	1·281	Unfailed slopes, −0·990
Constant		−6·219	Failed slopes, 0·996
DF2†			
β	0·867	0·124	General, 88·8
A	0·637	1·170	Partial (unfailed–failed), 82·0–95·6
D	−0·442	−0·406	
TS	−0·259	−0·325	
γ	−0·234	−0·025	Group centroids
Z	0·238	1·495	Unfailed slopes, −0·998
Constant		−7·258	Failed slopes, 1·005

*Eigenvalue, 0·993; Canon. corr., 0·706; Wilks-λ, 0·501; χ^2, 189·4; Significance, <0·0001.
†Eigenvalue, 1·011; Canon. corr., 0·709; Wilks-λ, 0·497; χ^2, 191·4; Significance, <0·0001.

Positive discriminant coefficients are associated with failed slopes and negative ones with unfailed slopes. Therefore, high values of β, A and Z increase discriminant scores and, consequently, slope instability. High values of LU, CRS and γ with a negative coefficient, increase stability. According to this, failures are expected on non-forested steep slopes with large and gentle watershed, at locations showing concave transverse sections where colluvium may be accumulated. The discriminant function also expresses the ability of groundwater flow to reach the potential landslide site.

Discriminant scores range from 3·82 to −2·56. Failed and unfailed slope populations are characterized respectively by centroid values of $C_{\mathrm{f}} = 0{\cdot}996$ and $C_{\mathrm{un}} = -0{\cdot}990$. The distance between centroids shows that separation between the two groups by the discriminant function has been successfully achieved and that it can be used to assess shallow landslide susceptibility. By using a value of −0·4 as discriminant index (DI), 88·5 per cent of overall slope population was correctly classified, increasing up to 95·6 per cent when considering only failed slopes.

A new discriminant function ($FD2$) was obtained with the same variables using tree density (D) instead of land use (LU). The stepwise method gave nearly equal results, as can be seen in Table IV. The classification of slopes has improved by an insignificant 0·3 per cent. Therefore, either LU or D may be used without apparent changes in the discriminant function. However, given that the variable land use (LU) can be obtained in a much faster and cheaper way than tree density (D), it was finally selected to test the function.

A final analysis was performed including those variables that were first rejected because they did not adjust to a normal distribution: altitude and slope aspect. The aim of this trial was to determine the contribution and influence of these variables on slope stability. As stated already, moisture content of the soil will be affected by slope aspect. However, slope aspect may also reflect both the exposure to storm paths and vegetation cover. In the study area, southern oriented slopes are less covered with forest than northern ones. Furthermore, heavy rains that caused widespread landsliding activity in November 1982 came mainly from the south and southeast (Clotet and Gallart, 1983) and 59·8 per cent of landslides in the study area occurred on southern oriented slopes. On the other hand, the amount of rainfall during an intense storm event is expected to increase with altitude (Carrara *et al.*, 1978; Gallart and Clotet, 1988).

The results of the discriminant analysis are shown in Table V. The discriminant function (DF3) is composed of the same variables as used in previous functions with the inclusion of the altitude (H). However,

Table V. DF3 discriminant function results

Variables	Function coefficient		Correctly classified (%)
	Standard	Unstandard	(DI = −0·1)
β	0·952	0·141	General, 85·3
A	0·447	0·733	Partial (unfailed–failed), 78·9–91·6
LU	−0·440	−0·109	
TS	−0·234	−0·279	
γ	−0·243	−0·026	Group centroids
Z	0·372	2·214	Unfailed slopes, −1·014
H	0·427	0·002	Failed slopes, 1·003
Constant		−7·346	

Eigenvalue, 1·028; Canon. corr., 0·711; Wilks-λ, 0·493; χ^2, 131·2; Significance, <0·0001.

classification of the slopes has not improved. Conversely, the amount of slopes properly classified has been reduced by 3 to 4 per cent. Slope aspect was not significant enough to be included in the function.

TESTING OF THE DISCRIMINANT FUNCTION

The discriminant function has been applied to a pilot area in order to test its performance and reliability. Several procedures exist for testing landslide prediction models: (a) selection of a random sample to build the model and use of the remaining population to verify it (Neuland, 1976); (b) derivation of models from different random sample sizes and checking whether the function coefficients change significantly (Carrara, 1984); (c) preparation of the model from a distribution of landslides, which occurred during a specific event, and checking it with landslides triggered by a subsequent event (Luzi, 1995); and (d) development of the model in a training area, and testing it in a target area with similar characteristics. The last was used to validate our discriminant function.

The test area is located next to the Ensija range (Eastern Pyrenees), showing a variety of morphological and lithological features. The area was divided into cells of 50 m × 50 m. At each cell, the geologic and geomorphologic attributes (variables) used in the discriminant analysis were collected along with the indication of presence or absence of slope failure. Not all variables of the discriminant function were used. As thickness of superficial deposits (Z) is a variable difficult to obtain, unless natural cuts exist, it was excluded from the function. A new discriminant function was derived with the remaining variables (β, A, LU, CRS and γ). Table VI shows the new function (*DF4*) with its statistical controls and Figure 2 shows the histogram of frequencies of the same function. Every cell will take a value given by the discriminant function (discriminant score), which defines its landslide susceptibility level. The discriminant value of −0·3 was used for grouping the cells according to their susceptibility to landsliding. Cells with scores higher than −0·3 were considered as susceptible to producing slope failures. Correspondence between cells classified as susceptible and slope failures was 87·3 per cent.

Finally a landslide-susceptibility map was prepared using discriminant function *DF4*. Discriminant scores were divided into several ranges, based on the percentage of failed slopes with respect to the whole slope population. Five susceptibility classes were established according to these ranges (Figure 3):

I	(very low)	$DF < -1{\cdot}8$
II	(low)	$-1{\cdot}8 \leq DF < -0{\cdot}9$
III	(moderate)	$-0{\cdot}9 \leq DF < 0{\cdot}1$
IV	(high)	$0{\cdot}1 \leq DF < 1{\cdot}4$
V	(very high)	$DF \geq 1{\cdot}4$

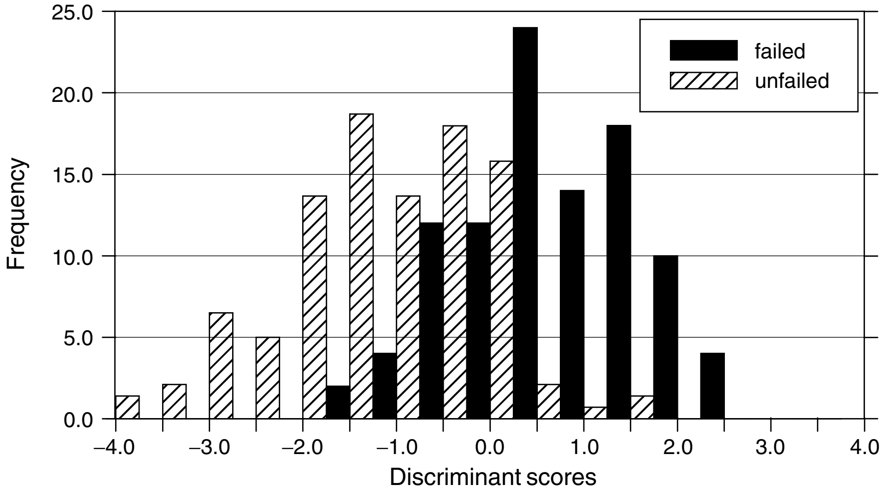

Figure 2. Frequency histogram of DF4 discriminant scores

Table VI. DF4 discriminant function results

Variables	Function coefficient		Correctly classified (%)
	Standard	Unstandard	(DI = −0·3)
β	0·875	0·125	General, 84·65
A	0·744	1·365	Partial (unfailed–failed), 74·3–95·0
LU	−0·396	−0·098	
TS	−0·245	−0·308	Group centroids
γ	−0·229	−0·025	Unfailed slopes, −0·974
Constant		−6·555	Failed slopes, 0·981

Eigenvalue, 0·962; Canon. corr., 0·700; Wilks-λ, 0·509; χ^2, 185·0; Significance, <0·0001.

Taking into account that the value of (−0·3) was used to differentiate susceptible and non-susceptible slopes, it may be expected that slope failures would appear in cells having higher discriminant scores (susceptibility levels III to V). In order to verify this, an index of relative landslide density has been used. This index is defined by the ratio between the density of slope failures of a given susceptibility class and the overall slope failure density. The index takes the following form: $100(n_i/N_i)/\Sigma(n_i/N_i)$, where n_i is the number of slope failures observed within a susceptibility class and N_i is the area occupied by the cells of this class. Table VII shows the relative landslide density for each susceptibility class. We may conclude that the distribution of slope failures observed in these classes indicates that susceptibility levels are consistent.

Table VII. Index of relative landslide density for DF4 discriminant function on pilot area

Susceptibility levels	Landslides (n = 55)	Cells (N = 986)	$100(n/N)/\Sigma(n/N)$
I, very low	0	243	0
II, low	2	297	1·61
III, moderate	15	205	17·56
IV, high	22	161	32·80
V, very high	16	80	48·01

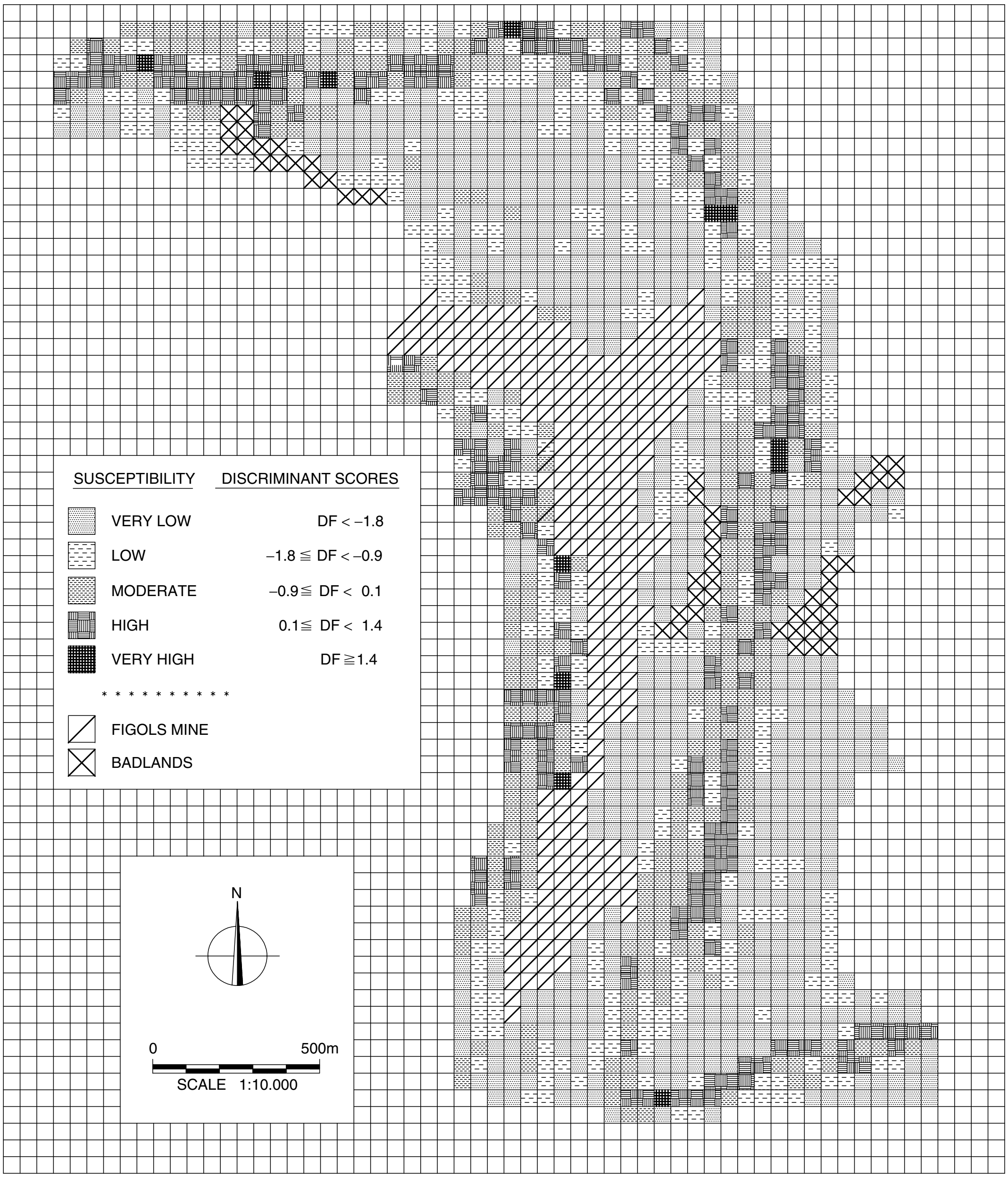

Figure 3. Shallow-landslide-susceptibility map by DF4 discriminant function

FINAL REMARKS

The spatial landslide distribution is the result of the interaction of many factors, some of which are difficult to incorporate in susceptibility analysis. A reliable and accurate hazard assessment depends on the proper identification of these factors. The inclusion or omission of some of the factors may change significantly the capability of susceptibility assessment, yet data are often gathered because of their availability rather than their suitability as predictors of slope stability conditions.

Among available methodologies for susceptibility zoning of large areas, statistical techniques have proved very useful. In this investigation we used multivariate analysis for the prediction of landslide susceptibility in the Eastern Pyrenees. The discriminant function obtained in this research allows the successful assessment of landslide susceptibility using only five factors that can be easily gathered in the field or with automatic data capture techniques.

The first two attempts showed a successful discrimination between stable and unstable slopes with results ranging from 88·8 per cent to 88·5 per cent of cases correctly classified, but always with an error of unstable slopes lower than 5 per cent. In both cases, the same variables were selected as the most influential factors on stability. High values of slope angle (β), watershed area (A) and thickness of superficial deposits (Z) increase slope instability, while land use (LU), cross-section (CRS) and mean watershed angle (γ) work in the opposite sense. Summing up, shallow landslide likelihood is higher on slopes showing steep angles, large unforested watersheds, and at locations showing concave transverse sections where colluvium may be accumulated. This discriminant function also expresses the capability of groundwater flow to reach the potential failure site.

The exclusion of soil type (ST), does not imply that lithology is not an important factor in slope stability, but is due to the homogeneity of the lithological formations in the studied area. Future work performed in areas with heterogeneous lithologies may demonstrate its influence on stability. On the other hand, the inclusion of altitude (H) in the discriminant function should be interpreted exclusively in terms of its high correlation with rainfall intensity distribution.

Finally, multivariate techniques provide quantified evaluation of the simultaneous influence of different factors and therefore a more realistic and objective approach to the assessment of landslide susceptibility.

ACKNOWLEDGEMENT

This research has been supported by the Spanish Scientific Council-CICYT (project AMB97-1091-C06).

REFERENCES

Anbalagan R. 1992. Landslide hazard evaluation and zonation mapping in mountainous terrain. *Engineering Geology* **32**: 269–277.

Baeza C. 1994. *Evaluación de las condiciones de rotura y de la movilidad de los deslizamientos superficiales mediante el uso de técnicas de análisis multivariante*. Doctoral thesis, Universidad Politécnica de Catalunya.

Bonham-Carter GF, Agterberg FP, Wright DF. 1990. Weights of evidence modelling: a new approach to mapping mineral potential. In *Statistical Applications in the Earth Sciences*, Agterberg FP, Bonham-Carter GF (eds). Geological Survey of Canada, Paper 8–9: 171–183.

Carrara A. 1983a. Multivariate models for landslide hazard evaluation. *Mathematical Geology* **15**(3): 403–426.

Carrara A. 1983b. Geomathematical assessment of regional landslide hazard. In *Proceedings of the 4th International Conference on the Applications of Statistics and Probability in Soil and Structural Engineering*. Augusti G, Vannucchi G (eds). (Università di Firenze) Pitagora Editrice, Bologna; 3–27.

Carrara A. 1984. Landslide hazard mapping: aims and methods. *Bureau de Recherches Gèologiques et Minière* **83**: 141–151.

Carrara A, Catalano E, Sorriso-Valvo M, Realli C, Ossi I. 1978. Digital terrain analysis for land evaluation. *Geologia Applicata e Idrogeologia* **13**: 69–127.

Chung CF, Fabbri A. 1993. The representation of geoscience information for data integration. *Nonrenewable Resources* **2**(2): 122–139.

Chung CF, Leclerc Y. 1994. A quantitative technique for zoning landslide hazard. In *International Association for Mathematical Geology Annual Conference* (IAMG'94) Quebec, Canada: 87–93.

Clotet N, Gallart F. 1983. *Inventari de degradacions de vessants originades pels aiguats de Novembre de 1982, a les altes conques del Llobregat i Cardener* Serv. Geoll. Generalitat de Catalunya; Institut Jaume Almera CSIC.

Corominas J. 1992. Landslide risk assessment and zoning. In *Planning the Use of the Earth's Surface*, Cendrero A, Lütting G, Wolf FC (eds). Springer-Verlag: Berlin; 141–173.

Corominas J, Alonso E. 1990. Geomorphological effects of extreme floods (November 1982) in the Southern Pyrenees. In *Proceedings of the 2nd Symposium of Hydrology in Mountainous Regions Artificial Reservoirs: Water and Slopes*. International Association of Hydrological Sciences, Lausanne, **194**; 295–302.

Corominas J, Baeza C, Salueña I. 1992. The influence of geometrical slope characteristics and land use on the development of shallow landslides. In *Proceedings of the 6th International Symposium on Landslides* D. H. Bell (ed.). 919–924.

Corominas J, Moya J. 1999. Reconstructing recent landslide activity in relation to rainfall in the Llobregat River basin, Eastern Pyrenees, Spain. *Geomorphology* **30**: 73–93.

David M, Dagbert M, Beaucheimin Y. 1977. Statistical analysis in geology: Correspondence analysis method. *Quarterly Journal of the Colorado School of Mines* **72**: 1–45.

Dillon WR, Goldstein M. 1986. *Multivariate Analysis. Methods and Applications*. John Wiley: Chichester.

Gallart F, Clotet N. 1988. Some aspects of the geomorphic processes triggered by an extreme rainfall event: The November 1982 flood in The Eastern Pyrenees. *Catena supplement* **13**: 79–95.

Gao J. 1993. Identification of topographic setting conducive to landsliding from DEM in Nelson County, Virginia (USA). *Earth Surface Processes and Landforms* **8**: 579–591.

Greenway DR. 1987. Vegetation and slope stability. In *Slope Stability*, Anderson MG, Richards KS (eds). John Wiley: Chichester; 187–231.

Hansen MJ. 1984. Strategies for classification of landslides. In *Slope Instability*, Brunsden D (ed.). John Wiley: Chichester; 1–25.

Hartlen J, Viberg L. 1988. Evaluation of landslide hazard. In *Proceedings of 5th International Symposium on Landslides*. Bonnard CH (ed.). Balkema: Rotterdam; **2**: 1037–1057.

Johnson KA, Sitar N. 1989. Hydrologic conditions leading to debris-flow initiation. *Canadian Geotechnical Journal* **27**: 789–801.

Jones FO, Embody DR, Peterson WL. 1961. *Landslide along the Columbia river valley, Northeastern Washington* US Geological Survey Professional Paper 367.

Kawakami H, Saito Y. 1984. Landslide risk mapping by a quantification method. In *Proceedings of 4th International Symposium on Landslides*. University of Toronto: Downsview; **2**: 535–540.

Lebart L, Morineau A, Fenelan JP. 1982. *Traitement des donées statistiques*. Dunad: Paris.

Luzi L. 1995. *GIS for slope stability zonation in the Fabriano area, central Italy* PhD thesis, International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede.

Mora S, Vahrson WG. 1994. Macrozonation methodology for landslide hazard determination. *Bulletin of the Association of Engineering Geologists* **XXXI**(1): 49–58.

Mulder HF. 1991. *Assessment of Landslide Hazard* Faculty of Geographical Science, University of Utrecht.

Muñoz JA. 1985. *Estructura alpina i herciniana a la vora sud de la zona axial del Pirineu Oriental* Doctoral thesis Facultad de Geologia, Universitat de Barcelona.

Muñoz JA, Martínez A, Vergés J. 1986. Thrust sequences in the Spanish Eastern Pyrenees. *Journal of Structural Geology* **8**(3/4): 399–405.

Neuland H. 1976. A prediction model of landslips. *Catena* **3**: 215–230.

Nie NH, Hull CH, Jenkens JG, Steinbrenner K, Bend OM. 1981. *Statistical package for the social sciences*. Suppl. 2. McGraw-Hill: New York.

Nilsen TH, Wright RH, Vlasic TC, Spangle VE. 1979. *Relative slope stability and land-use planning in the San Francisco Bay Region, California* US Geological Survey Professional Paper 944.

Novoa M. 1984. Precipitaciones y avenidas extraordinarias en Catalunya. *Jornadas de trabajo sobre Inestabilidad de laderas en el Pirineo* **I**: 1.1–5.1.

Okimura T. 1983. A prediction system for the site of probable surface failure of mountain-slope by topographical factors. *Proc. Japan Society of Civil Engineers*; **338**: 131–138.

Oyagi N. 1984. Landslides in weathered rocks and residual soils in Japan and surrounding areas: A state of the art report. In *Proceedings of the 4th International Symposium on Landslides*. University of Toronto: Downsview; **3**: 1–31.

Puigdefábregas C, Serrat D, Vilaplana JM. 1979. El medi geològic. In *Patrimoni natural d'Andorra*, Folch R (ed.). Ketres editorial: Barcelona; 15–39.

Smith TC. 1988. A method for mapping relative susceptibility to debris flow, with an example from San Mateo County. In *Landslides, Floods and Marine Effects of the Storm of January 3–5, 1982, in San Francisco Bay Region, California*, US Geological Survey: United States; Professional Paper; **1434**: 185–194.

SPSS Inc. 1988. *SPSS-X Users Guide*, 3rd edn. Gorinchen.

Van Westen CJ. 1994. GIS in landslide hazard zonation: a review with examples from the Andes of Colombia. In *Mountain Environment and Geographic System*. Price M, Heywood DI (eds). Taylor and Francis: London; 135–165.

Yin KL, Yan TZ. 1988. Statistical prediction models for slope instability of metamorphosed rocks. In *Proceedings of the 5th International Symposium on Landslides*, Bonnard CH (ed.). Balkema: Rotterdam; **2**: 1269–1272.

Zaruba Q, Mencl V. 1969. *Landslides and Their Control*. Elsevier: Amsterdam.

