

Statistical and spatial analysis of landslide susceptibility maps with different classification systems

Cristina Baeza¹ · Nieves Lantada²  · Samuel Amorim³

Received: 17 May 2016 / Accepted: 23 September 2016 / Published online: 3 October 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract A landslide susceptibility map is an essential tool for land-use spatial planning and management in mountain areas. However, a classification system used for readability determines the final appearance of the map and may therefore influence the decision-making tasks adopted. The present paper addresses the spatial comparison and the accuracy assessment of some well-known classification methods applied to a susceptibility map that was based on a discriminant statistical model in an area in the Eastern Pyrenees. A number of statistical approaches (Spearman's correlation, kappa index, factorial and cluster analyses and landslide density index) for map comparison were performed to quantify the information provided by the usual image analysis. The results showed the reliability and consistency of the kappa index against Spearman's correlation as accuracy measures to assess the spatial agreement between maps. Inferential tests between unweighted and linear weighted kappa results showed that all the maps were more reliable in classifying areas of highest susceptibility and less reliable in classifying areas of low to moderate susceptibility. The spatial variability detected and quantified by factorial and cluster analyses showed that the maps classified by quantile and natural break methods

were the closest whereas those classified by landslide percentage and equal interval methods displayed the greatest differences. The difference image analysis showed that the five classified maps only matched 9 % of the area. This area corresponded to the steeper slopes and the steeper watershed angle with forestless and sunny slopes at low altitudes. This means that the five maps coincide in identifying and classifying the most dangerous areas. The equal interval map overestimated the susceptibility of the study area, and the landslide percentage map was considered to be a very optimistic model. The spatial pattern of the quantile and natural break maps was very similar, but the latter was more consistent and predicted potential landslides more efficiently and reliably in the study area.

Keywords Landslides susceptibility maps · GIS · Kappa index · Discriminant analysis

Introduction

Landslide susceptibility maps (LSMs) are essential tools for land-use spatial planning in mountain environments. These maps are built with predictive models based on complex and sophisticated mathematical methods (discriminant, logistic regression, neural network, etc.) using large databases on landslide influencing parameters. The main reason for displaying the susceptibility data cartographically is to facilitate the spatial patterns taking care to depict as accurately as possible the underlying distribution of data (Cromley and Mrozinski 1997). An ideal classification system seeks to strike a balance between the underlying data and the simplification of the continuous susceptibility values that reveal intrinsic spatial patterns. There are different methods of classifying the susceptibility

✉ Nieves Lantada
nieves.lantada@upc.edu

¹ Scientific Department, ACUIDAD Consulting, Barcelona, Spain

² Division of Geotechnical Engineering and Geosciences, Department of Civil and Environmental Engineering, Universitat Politècnica de Catalunya, Jordi Girona 1-3, Building D2, 08034 Barcelona, Spain

³ Universidade Católica de Pernambuco (UNICAP), Recife, PE, Brazil

that enable us to simplify the information and to facilitate comprehension. These methods divide the susceptibility histogram, obtained from the map, into different classes. However, the classification systems have an inherent weakness, i.e., the aggregation of data in one of the classes may have an adverse effect on the apparent results depending on the criteria used when preparing maps. By altering the boundary between classes, very different-looking maps can be created (Evans 1977). Hence, the overall accuracy of the classified map determines the reliability of the data for any application (Liu et al. 2007). The consequences should be analyzed carefully since the classification can play a major role in decision-making tasks (Kiang 2003). Meaningful and consistent measures of map accuracy are necessary to evaluate the suitability of the map for their particular application. However, there are few published papers about landslide susceptibility assessment that provide information on the different classification methods. Measures should therefore be adopted to analyze the weakness of a particular classification strategy (Powell et al. 2004) or to compare two or more classification techniques (Foody 2004).

A number of approaches have been developed to test the accuracy of landslide spatial prediction maps (Congalton 1991; Stehman and Czaplewski 1998; Smits et al. 1999) that can be used to assess the classification maps. The “error matrix” (also known as confusion matrix, confusion table and contingency table) is central to most measures of thematic map accuracy (Story and Congalton 1986; Smits et al. 1999; Foody 2004; Powell et al. 2004; Gupta et al. 2008). Once a prediction threshold has been adopted by the mathematical model, the binary prediction (failed/unfailed) can be compared with the landslide observed. This allows the construction of the error or confusion matrix that shows the number of correctly and incorrectly predicted observations. But landslide susceptibility maps normally classify areas into different degrees of potential landslide, defining areas, for instance, as having very high (VHS), high (HS), moderate (MS), low (LS) and very low (VLS) landslide susceptibility, which are rendered on an ordinal scale. This approach has been often used to assess the concordance of binary maps. In this study, it has been used to investigate the match of the landslide susceptibility levels based on different classification systems. At it is clear from the Safeland Project comparison results, there is no standard classification system in Europe. The Safeland report of Work Package 2.1 about “Harmonization and development of procedures for quantifying landslide hazard” provides a detailed comparison of landslide mapping among several European countries (Safeland 2010).

In order to classify susceptibility data, common classification schemes provided for software package or created manually to generate classes can be used. However, when predicted values obtained from a mathematical model are

transposed to a map with classes by different classification systems, the predictive maps, originally with the same predictive power, do not have the same meaning. Hence, in order to shed light on this issue, a set of landslide susceptibility maps (LSMs) developed from the same dataset by discriminant analysis (Baeza and Corominas 2001) using different classification systems were built. These LSMs were compared and their similarities, differences, efficacy and consistency were assessed. Subsequently, the map that best matched the information in the study area was chosen. In addition to the confusion matrix, Spearman’s coefficient, the kappa statistic, landslide relative density index and an analysis of spatial image between maps were also performed to measure the classification agreement among the maps. The multivariate techniques such as factorial and cluster analyses proved very useful to evaluate the spatial proximity between maps, complementing and reinforcing the aforementioned approaches.

Study area

La Pobla de Lillet occupies an area of 40 Km² in the Eastern Pyrenees, Spain (Fig. 1).

The altitude above sea level, computed by a digital elevation model (DEM) of 15 m regular grid, ranges from 814 to 1.645 m. The maximum slope gradient is 65° with a mean value of 23.6°. Lithologies in the study area are composed of sandstones, limestones, marls and flysch formations from Devonian to middle Eocene. These geological formations belong to a series of east–west thrusts dipping toward the north (Muñoz et al. 1986). The landslide triggering factor in the region is rainfall. High-intensity rains of short duration triggered debris flows and shallow slides in November 1982 with rainfall reaching 340 mm in 48 h (Corominas and Alonso 1990; Corominas and Baeza 1992). Landslide distribution was controlled by lithology and the geomorphological and hydrological characteristics of the slopes. Most slope failures were developed on colluvial deposits and occasionally on underlying weathered clayey formations with a thickness not exceeding 1 m. They were attributed to steep forestless slopes, preferentially ranging between 30° and 35°. On slopes greater than 45°, the absence of failures was due to the rock formations. The contribution of water through catchment areas exceeded 1000 m³ with a mean angle ranging between 25° and 30° helped to generate many landslides in the study area. The significance of these parameters with respect to slope stability in the study area is discussed in Baeza and Corominas (2001) and Santacana et al. (2003). The failures considered in this study are shallow landslides with small mobilized volumes (less than 10,000 m³) and do not exceed two meters in depth.

Fig. 1 Geographical location and landslide inventory of the study area

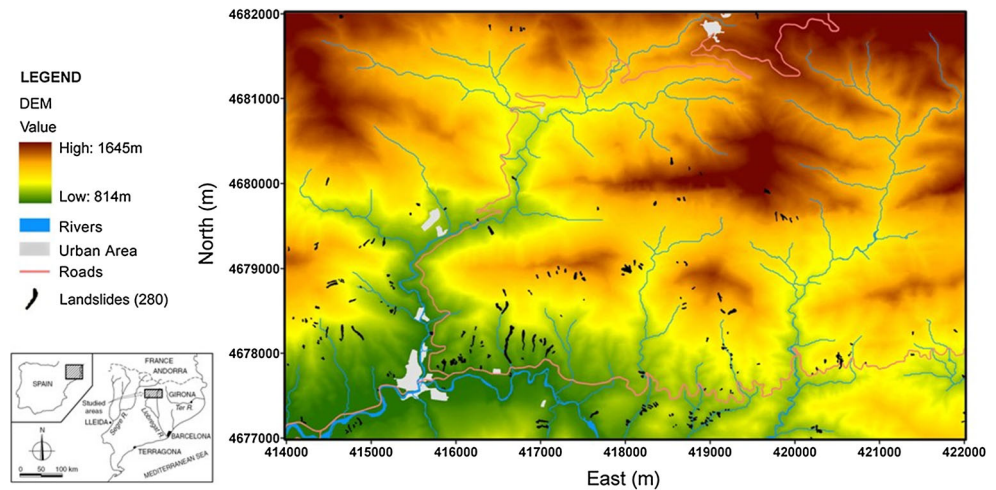


Table 1 Coefficients and statistical parameters of the discriminant model used to build the LSMs

Variables	Function coefficient		Statistics	
	Standard	Unstandard		
<i>Discriminant analysis</i>				
Height	-0.512	-0.003	Eigenvalue	0.652
Mean watershed angle	0.241	0.021	Canon. corr.	0.628
Sinusoidal slope angle	0.684	4.060	Wilks- λ	0.605
Slope aspect (0-180)	0.256	0.005	χ^2	140.282
Thickness of surf. dep.	0.405	0.556	Significance	0.000
Constant		-2.536		

Data source and landslide susceptibility map

The comparison and evaluation of different classification systems were carried out on a landslide susceptibility map generated by a discriminant model using the database by Santacana et al. (2003). This approach was used to assess the landslide susceptibility with parameters that provided indicators of the geomorphological evolution of the slope and valuable information for stability analysis. These parameters were derived directly from a digital elevation model (DEM) of 15 m regular grid supplied by the Cartographic Institute of Catalonia. The DEM was generated from a triangulated irregular network (TIN) using the topographic information at 1:5000 scale. Landslide inventory and thickness of the surficial formation were obtained from aerial photograph interpretation at 1:22,000, orthophotographs at 1:5,000 scale and field work. They were subsequently digitized and then converted into raster format for the analysis. The method of selecting these variables and their significance are discussed in earlier works by Baeza and Corominas (2001) and Hürlimann and Baeza (2002). The diagnosis, validation and model evaluation of the discriminant function obtained in the study area

are extensively described and analyzed in Baeza et al. (2010). Table 1 shows the variables and the main statistics of the discriminant function used to elaborate the LSMs with different classification systems.

There is a great deal of the literature on the different approaches that assess the susceptibility (e.g., logistic regression versus neural networks) or terrain units used in the mathematical model (e.g., pixel versus slope unit). All these approaches use classification systems to categorize the data in order to build the susceptibility map. The studies focus on the predicted susceptibility classification which translates into the reliability and accuracy of the models. However, in no case is the classification system applied to define the susceptibility levels analyzed. Classification involves a loss of information or a different redistribution of the data depending on the method applied. The articles by Coulson (1987) and Evans (1977) provide detailed overviews of the classification methods that are commonly employed. It is therefore important to determine the manner in which the classification affects the distribution of the susceptibility and its reliability with respect to other possible classification methods in the study area.

The method adopted in the literature to divide the susceptibility histogram into different categories is in many cases based on expert criteria (Dai and Lee 2002; Ohlmaier and Davis 2003; Van Den Eeckhaut et al. 2006) and does not take into account the real underlying data. It is necessary to explore data and obtain knowledge of their statistical distribution before applying any method of classification (Foote and Crum 2014). In this way, this categorization of data cannot be automated or statistically tested (Ayalew and Yamagishi 2005). At present, the cartographic representation in large regions needs an automatic and objective process. Automatic classification methods are increasingly being integrated into the geographical information system (GIS). Hence, in the present study, five classification methods were used to build five LSMs. These maps were divided into five categories (very low, low, moderate, high and very high). These categories were considered sufficient to reveal any existing spatial patterns in the dataset and facilitate map interpretation (Armstrong et al. 2003; Foote and Crum 2014).

Four of these classification systems are automatic and integrated in a GIS (ArcGis software):

Equal interval (EI) The range of susceptibility values is divided into equal-sized intervals.

Natural break (NB) Classes are based on natural groupings inherent in the data and boundaries are determined statistically where there are relatively large jumps in the susceptibility data values.

Quantile (Q) This is equivalent to equal coverage area, assigning the same number of cells in each class. In this case, the range of possible susceptibility values is divided into unequal-sized intervals. This classification scheme is well suited to linearly distributed data.

Standard deviation (SD) This shows the degree of deviation of pixel values from the mean; class breaks are then created using these values. Subsequently, adding or subtracting half deviation from the mean value of the data was used to define the susceptibility levels.

An additional classification method is a user-defined or manual classification system employing expert criteria:

Landslide percentage (LP) Based on the percentage of observed landslides in the area, as in the case of the aforementioned systems, five levels of landslide susceptibility were defined but in this case in accordance with the percentage of predefined landslides: very low (<1 % landslides), low (1–5 %), moderate (5–15 %), high (15–30 %), and very high (>30 %).

The resulting landslide susceptibility maps (LSMs) are shown in Fig. 2.

Comparative analysis of classification maps and results

Only when the goodness of fitting data and the prediction of capability of the susceptibility model defined by the mathematical function are confirmed, can the cartographic representation of the landslide susceptibility be performed. If these conditions are not met, it makes no sense to build the maps. Note that the set of LSMs developed from the same dataset using different classification methods show a normal distribution of discriminant values with the same predictive ability in origin (Fig. 2). Thereafter, a quantitative comparative analysis of the five LSMs was carried out to complement the common visual evaluation of the maps. This comparative evaluation allows us to identify quantitatively which maps are similar and which best define the landslide susceptibility of the study area.

Different approaches were adopted to measure the classification agreement and compare the LSMs: Spearman's rank correlation coefficient; kappa index; factor and cluster analyses; landslide density analysis (R index); and rank difference analysis.

Spearman's rank correlation coefficient

First of all, a nonparametric measure of statistical dependence between ranked variables (normality and homogeneity of variance assumptions are not satisfied) was calculated (Corder and Foreman 2009). This coefficient was used as a first evaluation of the statistical significance of the difference in the susceptibility classification between each pair of LSMs. The coefficient ranged from -1 to $+1$. The closer the index is to $+1$ or -1 , the stronger the probable correlation. A perfect positive correlation is $+1$ and a perfect negative correlation is -1 . This correlation could translate into a similar overall classification between maps. Table 2a shows that all classification systems have high correlations with values exceeding 0.85. The EI map differs from the other maps showing lower values of correlation (0.85–0.88). The correlation between the remaining maps ranges from 0.92 to 0.95. Table 2b shows the correlation only between cells with landslides, yielding values lower than those of the all the data sample.

The foregoing results suggest a fairly strong relationship between the LSMs. However, this statistic does not take into account the spatial location of each cell, i.e., the number of cells classified as high susceptibility could be the same for both maps but they could be located in different areas of the map. This index should therefore only be taken as indicative of an overall agreement in relation to the classification area covering. As regards the landslide cells, the agreement between the maps may be inconsistent

Fig. 2 Landslide susceptibility maps with five susceptibility levels using different classification system: **a** equal interval—EI; **b** natural break—NB; **c** quantile—Q; **d** standard deviation—SD; **e** landslide percentage—LP

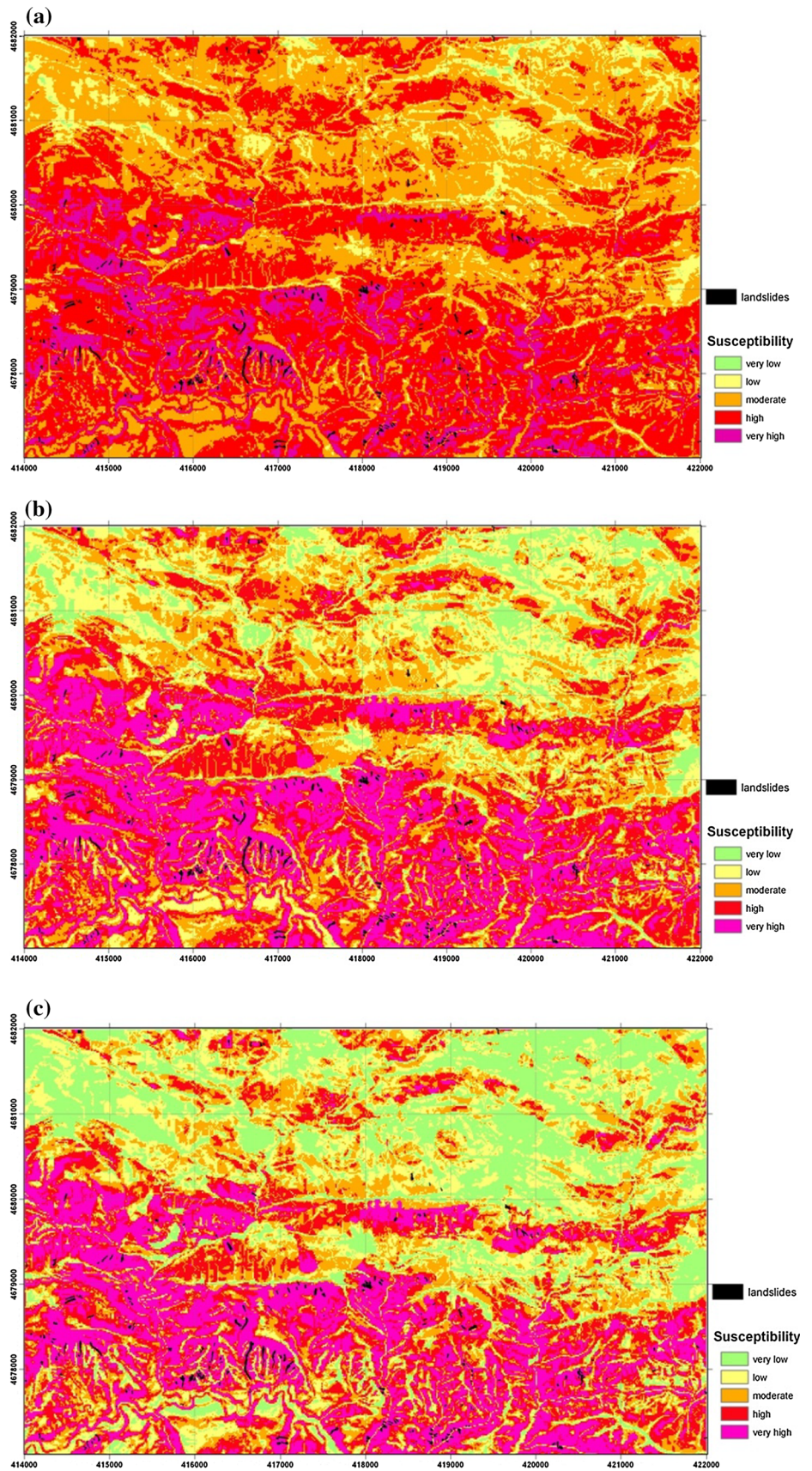


Fig. 2 continued

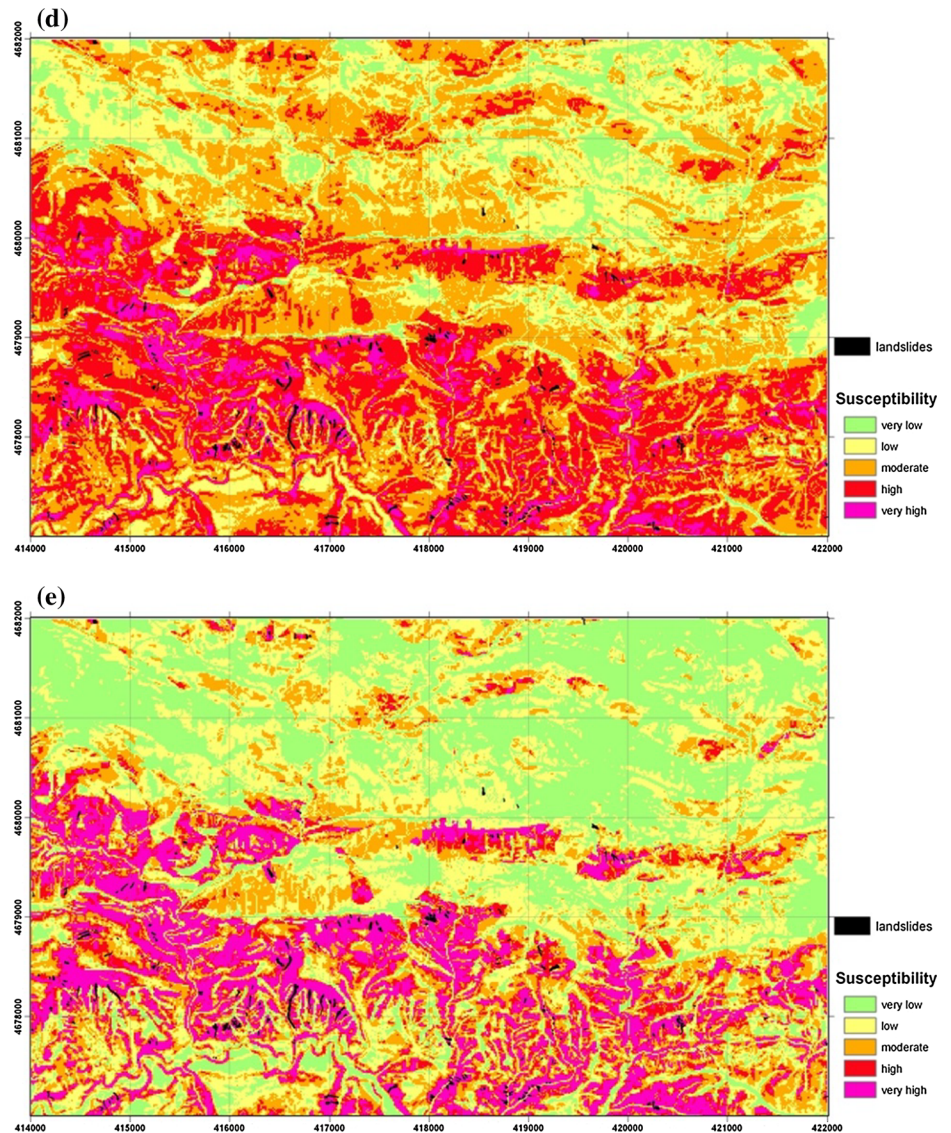


Table 2 Results of Spearman’s rank correlation between maps using (a) all the data and (b) only failed cells

	NB	Q	LP	Stand dev. (SD)
<i>(a) Spearman’s rank correlation coefficient (all the data)</i>				
Equal interval (EI)	0.888	0.888	0.859	0.855
Natural break (NB)		0.954	0.957	0.920
Quantile (Q)			0.954	0.922
Landslide percentage (LP)				0.940
<i>(b) Spearman’s rank correlation coefficient (failed cells)</i>				
Equal interval (EI)	0.557	0.573	0.711	0.789
Natural break (NB)		0.968	0.824	0.543
Quantile (Q)			0.842	0.555
Landslide percentage (LP)				0.634

Significance level 0.01

because the sample size of failed cells (270) is much lower than that of unfailed cells (177,362). In this type of statistic, the reliability of the results is related to sample size. The greater the amount of data collected, the more reliable the results (Poli and Sterlacchini 2007).

Unweighted and linear weighted kappa index

The kappa index was calculated to complete the information obtained by Spearman’s coefficient. This statistical index is considered a more robust measure than a simple observed proportion of agreement calculation since it also takes into account the proportion of agreement expected by chance. Kappa has a range from -1 to $+1$ with larger values indicating better concordance. The kappa statistic

Table 3 Unweighted kappa (*K*) and kappa with linear weighting (*K_{LW}*) between LSMs for (a) all the data and (b) only failed cells

	NB	Q	LP	Stand dev. (SD)
<i>(a) Unweighted kappa (K) and kappa with linear weighting (K_{LW}) for pairs of tests (all the data)</i>				
Equal interval (EI)	K = 0.41	K = 0.19	K = 0.04	K = 0.23
	se = 0.002	se = 0.001	se = 0.001	se = 0.002
	K_{LW} = 0.61	K_{LW} = 0.44	K _{LW} = 0.27	K_{LW} = 0.48
	se = 0.001	se = 0.001	se = 0.001	se = 0.001
Natural break (NB)		K = 0.57	K = 0.08	K = 0.57
		se = 0.001	se = 0.001	se = 0.002
		K_{LW} = 0.77	K_{LW} = 0.49	K_{LW} = 0.74
		se = 0.001	se = 0.001	se = 0.001
Quantile (Q)			K = 0.40	K = 0.47
			se = 0.002	se = 0.002
			K_{LW} = 0.70	K_{LW} = 0.70
		se = 0.001	se = 0.001	se = 0.001
Landslide percentage (LP)				K = 0.16
				se = 0.001
				K_{LW} = 0.53
			se = 0.001	se = 0.001
<i>(b) Unweighted kappa (K) and kappa with linear weighting (K_{LW}) for pairs of tests (failed cells)</i>				
Equal interval (EI)	K = 0.41	K = 0.39	K = 0.43	K = 0.66
	se = 0.055	se = 0.055	se = 0.051	se = 0.043
	K_{LW} = 0.46	K_{LW} = 0.46	K_{LW} = 0.51	K_{LW} = 0.69
	se = 0.047	se = 0.042	se = 0.030	se = 0.035
Natural break (NB)		K = 0.88	K = 0.36	K = 0.25
		se = 0.036	se = 0.066	se = 0.051
		K_{LW} = 0.90	K_{LW} = 0.56	K _{LW} = 0.35
		se = 0.023	se = 0.029	se = 0.042
Quantile (Q)			K = 0.43	K = 0.30
			se = 0.062	se = 0.050
			K_{LW} = 0.64	K_{LW} = 0.41
		se = 0.032	se = 0.045	se = 0.045
Landslide percentage (LP)				K = 0.33
				se = 0.047
				K_{LW} = 0.48
			se = 0.035	se = 0.035

Significance level 0.05

Numbers in bold express more than moderate agreement by Landis and Koch (1977)

se standard error

can be expressed in the following conceptual terms (Landis and Koch 1977):

$$Kappa = \frac{(d - q)}{(N - q)} \tag{1}$$

where *d* (observed agreement) is the proportion of cells in agreement, *q* is the proportion of agreement expected by chance and *N* is the total observations.

To compute this index, the error matrix (Beguera 2006) was prepared for each pair of LSMs. The procedure provides a series of statistical tests and measures of association

for double-sorting tables that allow us to evaluate the statistical significance of the kappa value and therefore the similarity between maps.

First, an unweighted kappa was evaluated. This index is now widely used in the literature (Guzzetti et al. 2006; Van den Eeckhaut 2006; Thiery et al. 2007; Sterlacchini et al. 2008) for the comparison of susceptibility models. In Table 3a, most kappa (*K*) values below 0.4 suggest that the LSMs are not as similar as revealed by Spearman’s coefficients, with some exceptions. The standard error (*se*) is also tabulated as a reliability measure of the study.

Table 4 Statistical descriptives (minimum, maximum, mean and standard deviation values) for all and failed sample using unweighted and weighted kappa

Statistical descriptives ($N = 10$)				
	Min.	Max.	Mean	SD
Un. kappa (as)	0.04	0.57	0.3120	0.19663
Un. kappa (fs)	0.05	0.88	0.4440	0.18846
W. kappa (as)	0.27	0.77	0.5730	0.15889
W. kappa (fs)	0.35	0.90	0.5460	0.16071

as all sample, fs failed sample

According to the scale defined by Landis and Koch (1977), which qualitatively expresses the force of the agreement based on the kappa index, the classification of the LSMs for the whole area rarely reaches a “moderate” agreement (0.4–0.6). Note that the LP map with very low kappa values is different from other maps. The global match increases when only the failed cells are taken into account, as shown in Table 3b. LP improves the kappa values, while the SD map worsens the global index when failed cells are used. This seems reasonable since the rating system of LP takes into account only landslide frequency, while SD uses the standard deviation of all the data. In line with the results of the kappa index, the Q map and the NB map are the ones that show the best agreement (0.57) when all the data of the map are used. They reach almost perfect agreement (0.88) when only landslide data are employed.

In the previous analysis, all disagreement is treated equally as total disagreement. However, the levels of susceptibility are ordered—level 2 represents greater probability to fail than level 1, level 3 represents greater probability to fail than level 2, and so on. It is therefore important to take into account not only absolute concordances but also relative concordances (Cohen 1968). The use of either index (weighted or unweighted) may indicate a different efficiency and reliability of the susceptibility map and hence its role in hazard and risk management. Thus, when categories are ordered, it is recommendable to use *weighted kappa* and assign different weights to categories so that different levels of agreement can contribute to the value of kappa. Different weights can be used, but in the present study the kappa with linear weighting was calculated (Fleiss et al. 2003). The kappa value penalizes linearly the disagreement between maps from the smallest to the largest. Table 3a (all the data) shows an overall improvement in weighted kappa over the unweighted kappa although relationships between maps still have the same trend. Some LSMs reach “good” agreement with values higher than 0.7. The weighted kappa values with respect to the unweighted kappa values show an average increase of 185 % for all the data, while this is only 28 % for failed cell data.

In order to account for the difference statistically between weighted and unweighted kappa for each sample (all the data and failed data), spatial software for related sample was used. When comparing mean values and the standard deviation of each sample at a confidence level of 95 % (Table 4), a statistical significance level of 5 % was found between the mean weighted and unweighted kappa values for the whole sample, whereas (p value = 0.004) this was not the case (p value = 0.207) for the failed sample (Foody 2004). Accordingly, the disagreement between susceptibility classification levels (one or more ranks) for failed cells is less than that for the remaining cells. The type of kappa used for the sample of failed cells has less influence than using the sample of unfailed cells, i.e., the reliability of classification of the LSMs is greater for high susceptibility levels (failed cells). The reliability of unfailed cell classification is low. This means that all maps identify and delimit correctly the areas of highest susceptibility.

Factor and cluster analyses

The spatial location of the cell values in the map was not considered in earlier statistical approaches. In the present study, the detection of structure data provided by the factor analysis is used to analyze and plot their spatial distribution (Liu et al. 2007). Hence, examination of the underlying relationships between the different LSMs defined as variables with five susceptibility classes—very low, low, medium, high and very high—allows us to quantify the spatial similarity between the maps.

Two analyses were conducted, one with complete data of the map (failed and unfailed cells) and another one with landslides (failed cells). The scores of the rotated component matrix (Varimax procedure) and their graphical representation on the scatter plot are shown in Table 5 and Fig. 3a, respectively. The initial variance explained for the first two components with respect to the total variance in all variables is higher than 93 % in both analyses. This value reflects a close similarity between the LSMs. Using the complete data (whole cells) of the map, the first component correlates more with NB, Q and LP and the second component with EI. SD with lower and similar values in both components becomes independent when a third component is extracted. The same structure of the LSMs arises when cells with only landslide (failed cells) are analyzed. SD is prominent and is much closer to EI in the second component. As regards the spatial rotated plots, LP and EI are the maps that have the greatest differences and Q and NB are the closest. SD behaves like an unstable variable depending on the sample. SD is closer to EI when only landslides are analyzed and closer to Q and NB when the complete map is analyzed.

Table 5 Rotated factor matrix with weights over 0.7 in bold for all the data and failed cells

Component	All the data			Failed cells		
	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)
<i>Initial eigenvalues</i>						
1	4.588	91.769	91.769	4.023	80.469	80.469
2	0.195	3.897	95.666	0.632	12.633	93.101
		1	2		1	2
<i>Rotated component matrix</i>						
EI		0.475	0.871		0.323	0.899
NB		0.757	0.616		0.925	0.332
Q		0.813	0.545		0.916	0.355
SD		0.696	0.662		0.406	0.848
LP		0.877	0.444		0.841	0.468

Numbers in bold express factor loadings over 0.7

A hierarchical cluster analysis (HCA) is an exploratory tool (Sterlacchini et al. 2008) that reveals natural groupings within a dataset. This analysis complements the results of the factor analysis, providing a classification tree which links the most similar LSMs progressively until all the maps are joined. Using the nearest neighbor as the cluster method, a hierarchical graph of the cluster solution in a dendrogram form is shown in Fig. 3b.

Maps are listed along the left vertical axis, and the horizontal axis shows the distance between LSMs when they are joined. In both datasets (all the data and landslides), the HCA confirms the factorial results. When all the data (failed and unfailed cells) are analyzed, the first cluster consists of NB and SD followed by Q which has the smallest distance. Subsequently, EI is joined to the cluster. Given the distance of LP from the junction, it is this map that presents the biggest difference with respect to the other maps. As regards failed cells, NB and Q are the closest. Another cluster with EI and SD is created at a considerable distance from the first cluster. LP finally joins the first cluster (NB, Q). Hence, two different groups of LSMs appear in the same way as landslide density was also analyzed by the “relative landslide density index *R*” (Baeza and Corominas 2001) defined as follows:

$$R = \left[\frac{n_i/N_i}{\sum n_i/N_i} \right] \cdot 100 \tag{2}$$

where n_i is the number of cells with failures within a susceptibility level and N_i is the total number of cells of this level. It may therefore be expected that slope failures will appear in cells with higher discriminant scores (from moderate to extremely high susceptibility levels).

The landslide frequency for each susceptibility level is displayed in Fig. 4a. The frequency reaches 92 % of landslides classified into very high and high levels for all LSMs except for LP with 84 %. The results of LP are

obviously different from the others because susceptibility levels were manually predefined. Only NB and Q attain 80 % of landslides classified in a very high level. However, these values vary considerably when frequency is evaluated with respect to the coverage area (Fig. 4b) for each level. The R index distribution (Fig. 4c) for the LSMs displays a progressive increase, concentrating mainly on the highest susceptibility level. The R index of the different susceptibility levels proved to be fairly similar for NB, EI and Q (81–85 %) with lower values for SD and LP (66–76 %). The distribution of the cells with failures (landslides) in these levels indicates that the susceptibility levels are more consistent using the natural break, equal interval and quantile classification systems.

Image analysis

The close agreement displayed by Spearman’s coefficient substantially decreased when the proportion of agreement expected by chance (kappa index) was taken into account. The spatial structure by factorial and cluster analyses confirmed the dissimilarities between the maps. Despite the fact that these statistical approaches allow us to determine quantitatively which LSMs are the closest, the visualization of the spatial location of the differences and similarities is not possible. The spatial distribution of the susceptibility levels in each cell enables us to determine the most accurate map with landslides and where the maps match. This would allow us to delimit the spatial risk better.

LSMs in Fig. 2 show a well-defined pattern for the distribution of the susceptibility zones. All the maps reflect a horizontal zoning, influenced by the geological structure, which divides the area into two susceptibility zones: north and south. The south zone is basically more susceptible than the north zone, and the visual differences between the LSMs are mainly restricted to the distribution of the

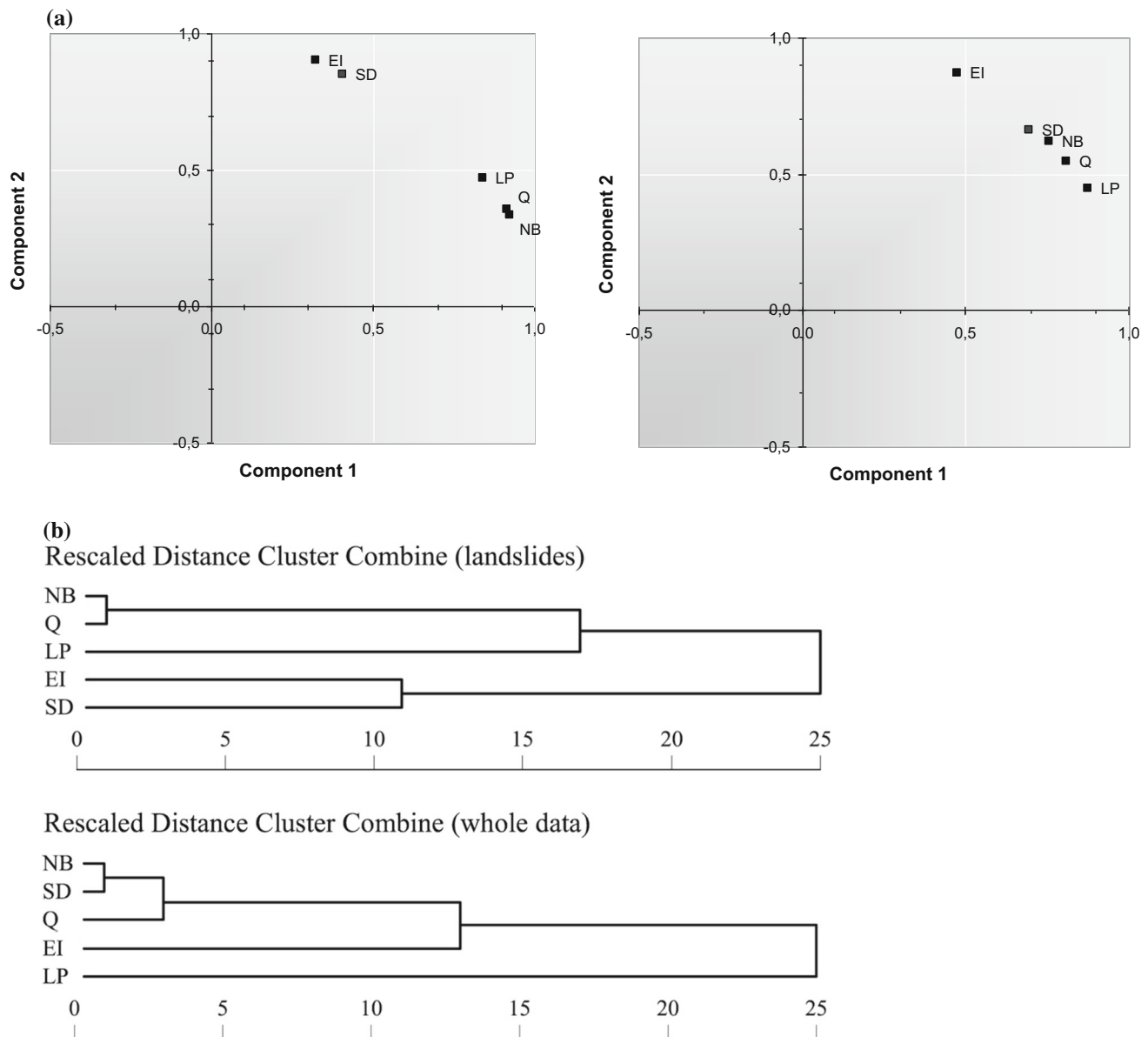


Fig. 3 **a** Spatial representation in two components by factorial analysis of the LSMs: using failed and unfailed sample (on the *left*) and using only failed sample (on the *right*). **b** Results of the hierarchical cluster analysis for failed sample (landslides) and whole sample

susceptibility levels in these two zones. A visual analysis shows a marked increase in the coverage area assigned as a very low susceptibility level from EI (0.4 %), NB (7.2 %), SD (7.5 %), Q (20.1 %) to LP (28.9 %) maps in the north zone. The high susceptibility in the South area is, however, restricted. It is therefore possible to refer to EI as the most conservative or pessimistic model and to LP as the liberal or optimistic model. The remaining models (NB, SD, Q) reveal intermediate trends.

In order to visualize the spatial match and mismatch of the five maps at a stroke, a procedure was implemented with GIS, extracting the susceptibility value of each cell. As a result, Fig. 5 displays the areas where some LSMs are

in agreement with the susceptibility level classification. It shows that the overall agreement between all LSMs is only 9.1 % of the coverage area. When only failed cells are considered, the agreement reaches 44.5 % of the covered area.

Analyzing the variables that define the discriminant prediction function, the mean values are higher in the areas where LSMs match than in the areas where they do not match with the exception of the *height* variable. LSMs agree in areas with higher slopes ($\bar{X}_{\text{match}} = 57^\circ$ vs. $\bar{X}_{\text{notmatch}} = 39^\circ$), higher watershed angles ($\bar{X}_{\text{match}} = 27^\circ$ vs. $\bar{X}_{\text{notmatch}} = 15^\circ$) and south-facing slopes ($\bar{X}_{\text{match}} = 123$ vs. $\bar{X}_{\text{notmatch}} = 98$) in lower elevation areas ($\bar{X}_{\text{match}} = 1073$ vs.

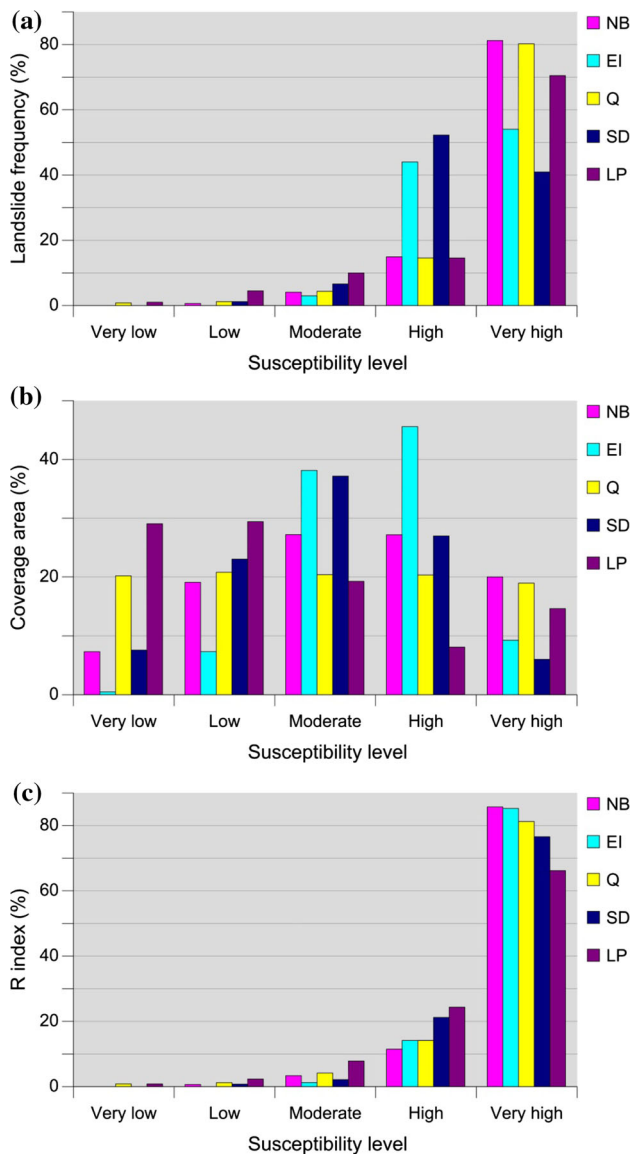


Fig. 4 Landslide frequency (a); coverage area (b) and R index (c) for the NB, EI, Q, SD and LP susceptibility maps

$\bar{X}_{\text{notmatch}} = 1237$ m). This means that the agreement between the maps is primarily in very high susceptibility levels and that they differ in classifying very low, low, and moderate susceptibility areas as shown in Fig. 6. In this figure, the mean values of the continuous variable of the discriminant function for each LSM are displayed. The figure clearly shows the disagreement between the maps from low to moderate susceptibility levels. For each variable, EI has the lowest mean values and LP the highest ones for very low, low, and moderate levels. As regards the *height* variable, EI and LP behave in an inverse way to that explained above. EI has the highest mean values and LP the lowest for these levels. This figure again illustrates the very conservative nature of EI versus the liberal LP. As for

the intermediate trends, there is agreement between NB and SD for low to moderate levels and between NB and Q for high and very high levels.

Despite the disagreement between the maps, it should be noted that the maximum rank of differences between LSMs is two susceptibility levels (red color), which accounts for 33.9 % of the area largely in the north zone (Fig. 7). More than half of the map (57 %) differs only in one level of susceptibility. This difference mainly concerns the central and southern areas of the map. Although this area is the most susceptible to failure, it is where the LSMs differ the least in the susceptibility level classification.

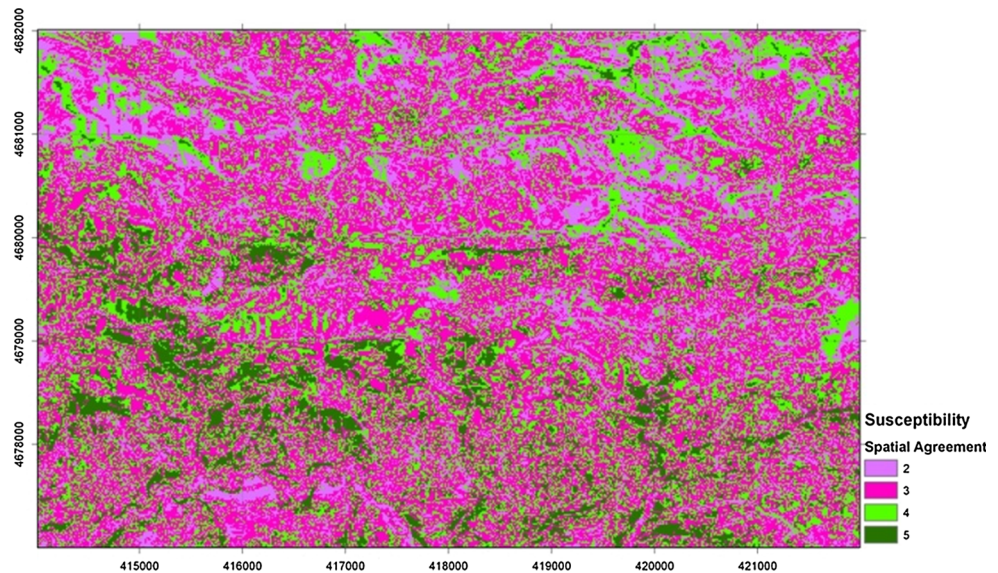
The map in Fig. 8 shows areas where the models agree. Only combinations of LSMs that are represented by more than 9 % of the area are displayed. Note that in this figure, when four LSMs are in agreement, LP is removed. Thus, LP is the map that differs most from the other maps, whereas NB and SD always appear together in all combinations, indicating their similarity in spatial classification.

In the light of the above results, EI and LP were removed from the following analysis because they did not adequately reflect the reality of the landslide susceptibility of the study area. EI shows a very pessimistic character, overestimating the susceptibility of the area. The opposite happens with LP, which is considered to be a very optimistic model, underestimating the susceptibility of a large part of the area. EI and LP are the two extreme classification models of the five LSMs analyzed.

Therefore, the NB, SD and Q susceptibility maps were compared in detail by difference image analysis. The maps were generated by subtracting the cell value (susceptibility level) of one LSM from the other. Thus, the final image can display the spatial distribution of the maximum difference between maps (Fig. 9). This figure also shows the distribution of the highest value of susceptibility level for each map in the area. There is only one level of difference between these maps.

Figure 9a displays the difference between NB and Q. The graph shows the agreement differentiating between unfailed and failed cells. The disagreement between cell types is evident. The percentage of unfailed cells that match is 65.2 %, whereas that of failed cells is 95.9 %, classifying most of the latter at the highest susceptibility levels (NB 95.9 %; Q 94.4 % see Fig. 4). The maximum difference is, in any case, only one level. Light blue zones show the distribution of the highest susceptibility level for NB. An uneven distribution of the light blue zone is shown, concentrating on the northern area of the map. NB overestimates the susceptibility of this zone with respect to Q. This suggests that NB is the most pessimistic or conservative of the two models in this area. Both maps show greater disagreement in areas where there is a lower

Fig. 5 Spatial agreement of the five landslide susceptibility maps. Color legend indicates where two or more LSMs match



frequency of landslides, namely in areas with a lower susceptibility.

Figure 9b displays the rank difference between NB and SD. The distribution of the highest susceptibility levels is inverse to the one shown above. NB reveals a higher susceptibility in the southern area than SD. The two maps show agreement in 67.4 % of unfailed cells, but only in 55.6 % of failed cells. The latter disagreement is due to the landslide classification in Fig. 4a. NB classifies 81.11 % of landslides in very high susceptibility, whereas SD only 40.74 %. NB is more conservative in the southern area than SD. However, this conservative nature of NB is more realistic than SD. NB reflects greater instability through the landslide frequency in the south, which is not reflected in SD.

The rank difference between Q and SD in Fig. 9c shows a low matching of landslides susceptibility zones throughout the area. Only 58 % match and 42 % cells exhibit a difference of one rank. This one-rank difference is distributed as follows: 22 % of the area is higher susceptibility for SD and 20 % for Q. Although the distribution area is almost equal, the north–south pattern is also very marked here. The susceptibility in the northern area is overestimated by SD and the southern area by Q. Then, Q better reflects the most susceptible zone of the area.

Discussion

Based on Spearman's rank correlation coefficient, as a first nonparametric measure of statistic dependence between maps, the five LSMs proved to be very similar (0.85–0.95), with a small difference of EI with respect to the remaining maps. This correlation value was reduced

when chance was taken into account by the kappa statistic. Kappa was defined in both unweighted and weighted (linear) forms. Despite the fact that the former coefficient is more usual in the literature, the latter is more accurate because the susceptibility level is an ordinal variable. Although the agreement calculated by the weighted kappa values was clearly different for unfailed and failed cells (“good” and “almost perfect,” respectively), these data cannot be compared owing to the prevalence of the unfailed data over the failed data (the greater the agreement between the maps, the smaller the sample). To resolve this problem, an inferential test between unweighted and weighted kappa was very useful. The mean value showed significant differences when the unfailed sample (p value = 0.004) was analyzed but not in the case of the failed sample (p value = 0.207) for a fixed significance level of 5 % ($\alpha = 0.05$). These findings confirm that LSMs are more reliable in classifying areas of the highest susceptibility level than areas of low to moderate susceptibility level.

Factorial and cluster analyses were performed to display graphically the similarity by distance between LSMs. They enable us to group similar maps considering the variance of spatial data for each map. The rotated spatial plots showed the greatest differences between EI and LP when all the data were analyzed. Q and NB were the closest and SD was the most unstable map depending on the sample data. SD proved to be more conservative when classifying failed cells with the result that it approached EI. Hierarchical cluster analysis provides a classification tree that links the most similar LSMs progressively, confirming the factorial results.

Prior to the image analysis, an evaluation of the classification of the susceptibility level was carried out by

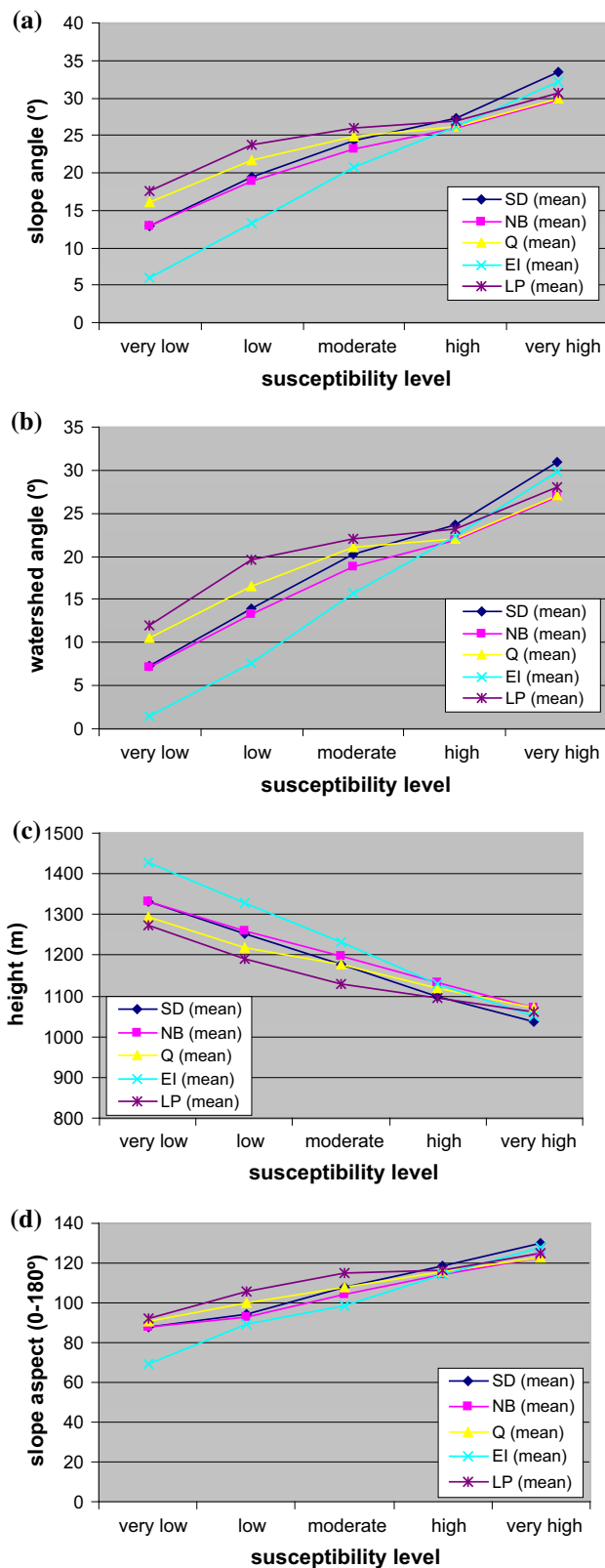


Fig. 6 Mean values of slope angle, watershed angle, height and slope aspect for each susceptibility level for the five LSMs

calculating the R index (landslide density analysis). The R index showed a consistent distribution given the greater frequency at the highest levels of the failed cells in the susceptibility levels for all LSMs. LP showed the worst classification followed by SD. As regards the coverage area for each level, the study area was more stable (very low and low susceptibility) for LP (58.08 %) than for the other maps. The most unstable model of the area was EI (7.58 %). If a model is biased in favor of safety, the above results confirm the optimistic or liberal nature of LP with respect to the pessimistic or conservative nature of EI in the study area.

Despite accuracy statistical measures, the analysis of image difference of LSMs is still very useful and highly revealing. All LSMs highlight the importance of the north-south orientation of the slopes owing to the geological structure (series of east-west thrusts) of the zone, showing a marked susceptibility pattern. However, the five LSMs only agree in 9 % of the area when classifying the susceptibility. This agreement is centered in the south-southwest where landslides are more frequent. These areas mainly correspond to levels of high and very high susceptibility. They are characterized by steeper slopes, steeper watershed angles, with bare slopes exposed to the sun at lower altitudes than areas where the LSMs disagree. They disagree basically in the northern area with susceptibility levels between very low and moderate. In these areas, EI is the most conservative map with the lowest mean values of susceptibility, whereas LP is the most liberal map with the highest values.

After rejecting the two most extreme models (one overestimates—EI—and the other underestimates—LP—the susceptibility) in the study area, intermediate models (NB, Q and SD) were compared by difference image analysis. This analysis is made by subtracting cell susceptibility value of one LSM from the other. The results show that the maximum difference between the three maps is only one level with an unequal distribution agreement. SD is furthest from the others with the highest disagreement in the failed cell classification (<58 %). This map underestimates the susceptibility in southern area with respect to NB and Q. On the other hand, the susceptibility in the south is reflected in a very similar way by NB and Q, reaching an agreement classification of 96 % for the failed cells. The disagreement is somewhat higher (65 %) for the remaining cells. This value is mainly in the north, where Q is more optimistic but not as realistic as NB. The reason for this is that Q does not consider the manner in which the data are distributed, thereby minimizing the intermediate susceptibility values of the function domain.

Fig. 7 Spatial distribution of the maximum difference of the susceptibility levels between the five LSMs

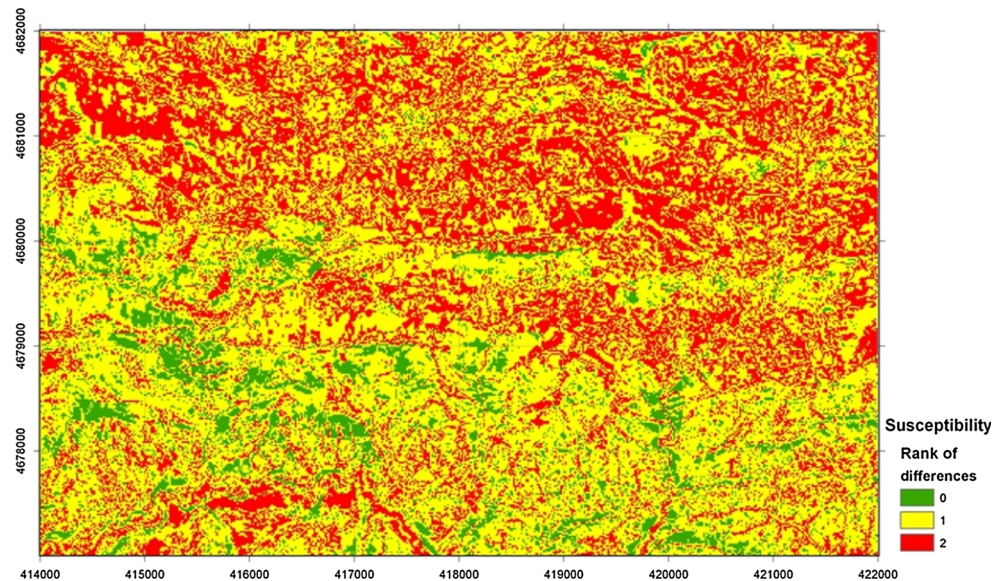
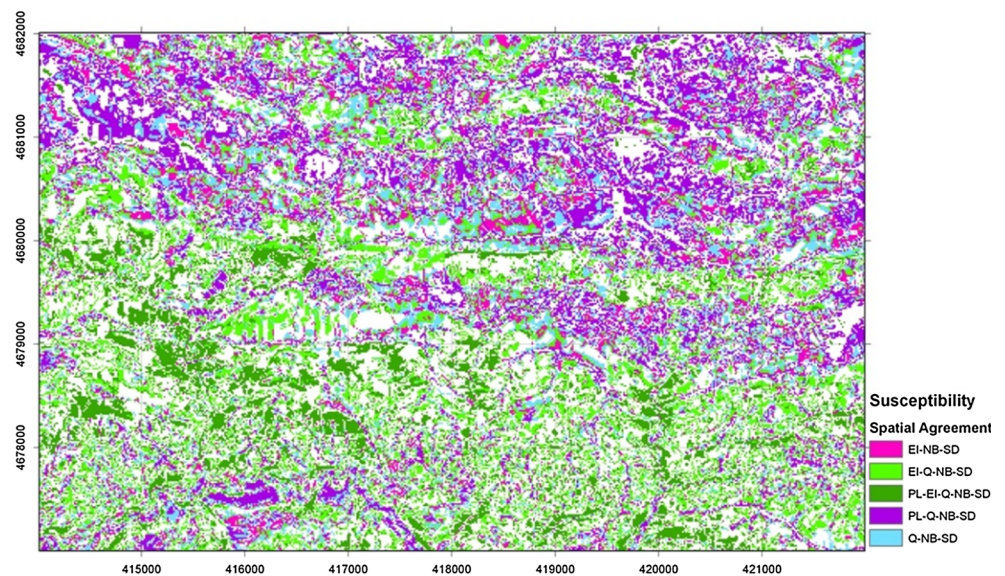


Fig. 8 Spatial agreement between some of the LSMs. Only the combinations higher than 9 % of the covered area have been displayed



Conclusions

Mapping landslide susceptible areas is essential for land-use spatial planning and management decision making in mountain environments. Landslide susceptibility is generated by mathematical models whose reliability must be confirmed. Then, classification systems are applied to predicted susceptibility values to obtain a landslide susceptibility map that is easy to interpret. However, the classification involves a loss of information that depends on the criteria adopted when building maps, which may seriously impair the apparent results. Hence, this study sought to compare statistically and rigorously five landslide susceptibility maps at La Pobla de Lillet (Spain) obtained by different classification systems (equal interval, natural

breaks, standard deviation, quantile and landslide percentage) in order to assess their similarities, differences, and their efficacy and consistency in the study area. The five maps ranked the predicted values of a discriminant model whose predictability and reliability had been confirmed in Baeza and Corominas (2001). A number of approaches (Spearman's correlation, kappa indexes, factorial and cluster analyses, landslide density index) to the comparison of map classification were used to complete and substantiate the usual image analysis of the maps.

The implementation of statistical measures consistent with the type of ordinal data (susceptibility levels) for analysis should be noted. Moreover, factors that can influence the magnitude of these measures (prevalence, bias and no independent ratings) were considered for a

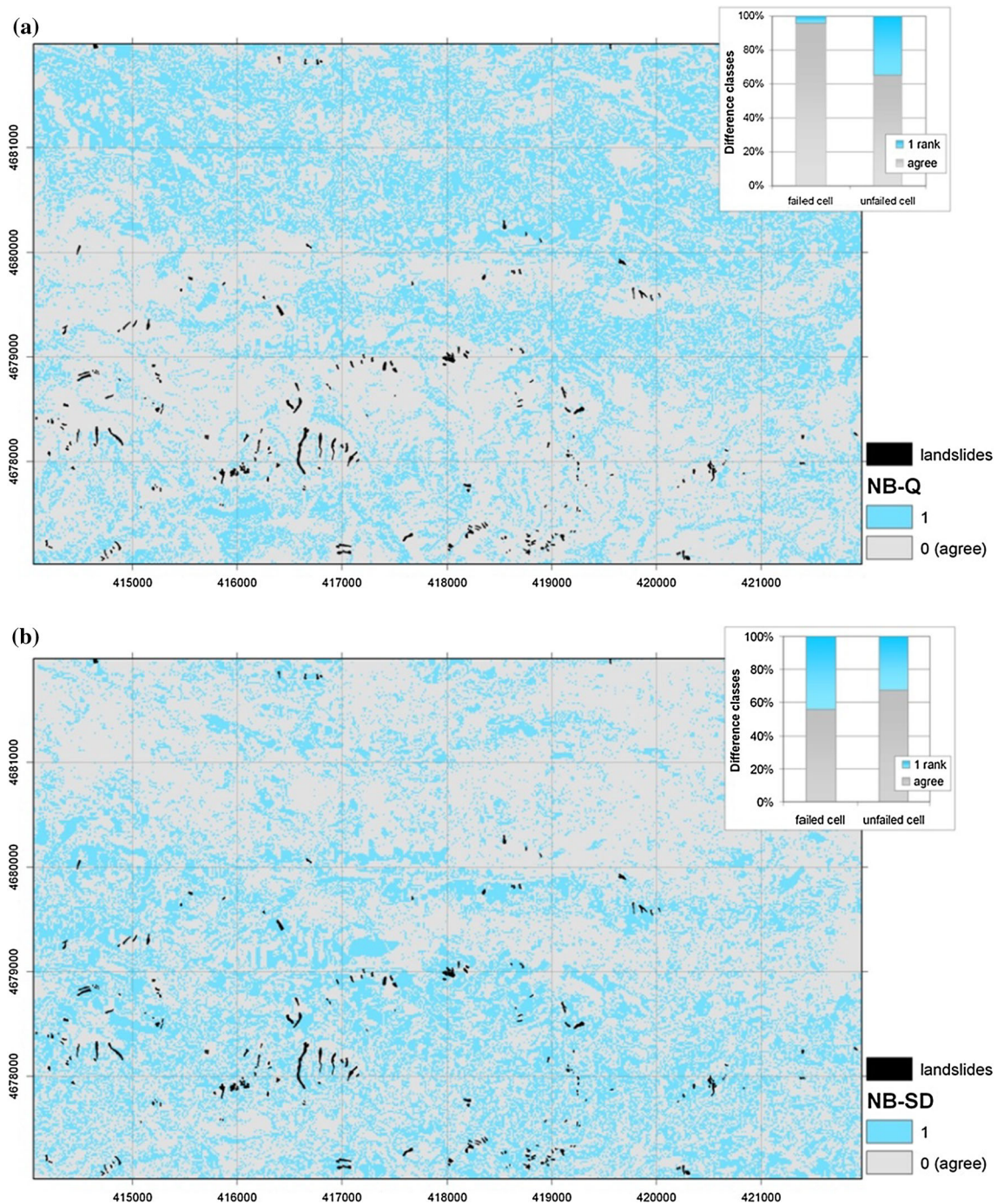


Fig. 9 Difference image between **a** NB and Q; **b** NB and SD; **c** Q and SD. Full matching cell (no difference) in susceptibility level in the two LSM is displayed in *gray*; one level difference can be displayed

in *light blue* or *dark blue*. *Light blue* zones correspond to higher susceptibility level for the first map in legend and *dark blue* zones (only in the Q and SD map) for the second map

correct interpretation of the results. Hence, nonparametric approaches with more statistical power provided measures of reliability in this study.

To sum up, the present study shows that despite using the same mathematical model with the identical prediction rate, the spatial agreement of these classification maps is not consistent and their spatial pattern is considerably

different. Thus, several statistical measures and spatial image analysis highlight the similarities and differences between the maps. The agreement between the maps was shown to be different. However, the accuracy of susceptibility levels increases only when the most susceptible areas are taken into account. This may be seen as a positive result, given that a high accuracy for the higher susceptible

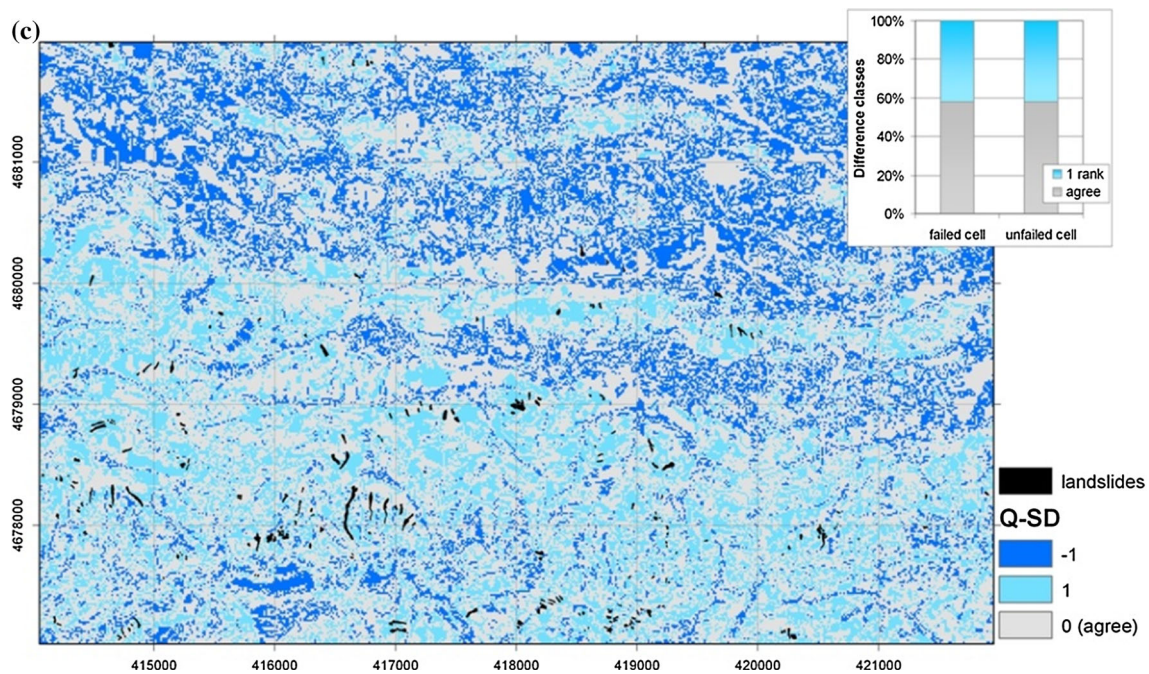


Fig. 9 continued

levels avoids the problem of identifying and classifying these dangerous areas. Notwithstanding, the over- or underestimation of the susceptibility in very low, low and moderate levels can have important implications for land management. The optimal map should be able to predict most of the potential landslides efficiently and reliably in the study area. Hence, the equal interval classification map (EI) could be regarded as excessively pessimistic (a large number of study cells are given high hazard susceptibility levels), while the landslide percentage map (LP) could be excessively optimistic (a large number of the study cells are given low hazard susceptibility levels). The former map may imply the loss of a potentially safe space, or even the uselessness of investments made for prevention in areas that could represent no danger. The latter map could lead to loss of life or the destruction of infrastructure as a result of incorrect classification of hazardous areas. LP is also a user-defined classification that is more difficult for the reader to interpret and is therefore harder to justify. Current automatic classification systems should therefore be used in place of a user-defined classification.

Of the three remaining LSMs, SD should be removed given that it does not achieve as good a landslide classification as other maps according to the R index. Finally, the spatial patterns of Q and NB are very similar, but Q is not as consistent as NB in relation to data distribution. As a result of this and given its easy implementation with respect to the other classification systems, NB is the most suitable classification map for modeling landslide susceptibility in the study area.

In the light of our findings, the particular classification strategy clearly determines the appearance of the landslide susceptibility map. The different maps obtained do not have the same meaning and may influence decision making. Hence, different classification systems should be analyzed and the one that best fits the structure of the landslide data in the study area should be adopted and adjusted to the needs of the end-user.

Acknowledgments The authors are indebted to the Cartographic and Geologic Institute of Catalonia (Spain) for providing the large-scale DEM of the study area.

References

- Armstrong MP, Xiao N, Bennett DA (2003) Using genetic algorithms to create multicriteria class intervals for choropleth maps. *Ann Assoc Am Geogr* 93(3):595–623. doi:[10.1111/1467-8306.9303005](https://doi.org/10.1111/1467-8306.9303005)
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65(1–2): 15–31. doi:[10.1016/j.geomorph.2004.06.010](https://doi.org/10.1016/j.geomorph.2004.06.010)
- Baeza C, Corominas J (2001) Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. *Earth Surf Proc Land* 26:1251–1263
- Baeza C, Lantada N, Moya J (2010) Validation and evaluation of two multivariate statistical models for predictive shallow landslide susceptibility mapping of the Eastern Pyrenees (Spain). *Environ Earth Sci* 61:507–523. doi:[10.1007/s12665-009-0361-5](https://doi.org/10.1007/s12665-009-0361-5)
- Begueria S (2006) Validation and evaluation of predictive models in hazard assessment and risk management. *Nat Hazards* 37(3): 315–329. doi:[10.1007/s11069-005-5182-6](https://doi.org/10.1007/s11069-005-5182-6)

- Cohen J (1968) Weighed Kappa: nominal scale agreement with provision for scale disagreement or partial credit. *Psychol Bull* 70(4):213–220. doi:10.1037/h0026256
- Congalton R (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens Environ* 37:35–46. doi:10.1016/0034-4257(91)90048-B
- Corder GW, Foreman DI (2009) *Nonparametric statistics for non-statisticians: a step-by-step approach*. Wiley, London. doi:10.1002/9781118165881
- Corominas J, Alonso E (1990) Geomorphological effects of extrem floods (November 1982) in the Southern Pyrenees. In: *Proceedings of 2nd symposia of hydrology in mountainous regions. Artificial reservoirs: water and slopes, Lausanne*. IAHS Publ. no 194. http://hydrologie.org/redbooks/a194/iahs_194_0295.pdf. Accessed 9 Sept 2016
- Corominas J, Baeza C (1992) Landslide occurrence in eastern pyrenees. *Movimenti franosi e metodi di stabilizzazione*. CNR 481:25–42
- Coulson MRC (1987) In the matter of class intervals for choropleth maps: with particular reference to the work of George F. Jenks. *Cartographica* 24(2):16–39. doi:10.3138/U7X0-1836-5715-3546
- Cromley RG, Mrozinski RD (1997) An evaluation of classification schemes based on the statistical versus the spatial structure properties of geographic distributions in choropleth mapping. 1997 ACSM/ASPRS Annual Convention & Exposition. Technical Papers, Seattle, Washington. <http://mapcontext.com/auto-carto/proceedings/auto-carto-13/pdf/auto-carto-13.pdf>. Accessed 5 Sept 2016
- Dai FC, Lee CF (2002) Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* 42(3–4):213–228. doi:10.1016/S0169-555X(01)00087-3
- Evans IS (1977) The selection of class intervals. *Trans Inst Br Geogr Contemp Cartogr* 2(1):98–124. doi:10.2307/622195
- Fleiss JL, Levin B, Paik MC (2003) *Statistical methods for rates and proportions*, book series: Wiley series in probability and statistics. Wiley, London. doi:10.1002/0471445428
- Foody GM (2004) Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogramm Eng Remote Sens* 70(5):627–633
- Foote KE, Crum S (2014) Issues of statistical generalization. *Cartographic communication*, section 6. Colorado, The Geographer's Craft Project, Department of Geography. The University of Colorado at Boulder. http://www.colorado.edu/geography/gcraft/notes/cartocom/cartocom_f.html. Accessed 5 Sept 2016
- Gupta RP, Kanungo DP, Arora MK, Sarkar S (2008) Approaches for comparative evaluation of raster GIS-based landslide susceptibility zonation maps. *Int J Appl Earth Obs Geoinf* 10(3):330–341. doi:10.1016/j.jag.2008.01.003
- Guzzetti F, Reichenbach P, Ardizzone F, Cardinali M, Galli M (2006) Estimating the quality of landslide susceptibility models. *Geomorphology* 81(1–2):166–184. doi:10.1016/j.geomorph.2006.04.007
- Hürlimann M, Baeza C (2002) Analysis of debris-flow events in the Eastern Pyrenees, Spain. In: *Proceedings of the first European conference on landslides, Prague, Czech Republic*. Eds. J. Rybar, Stemberk and Wagner. ISBN: 905809393X
- Kiang MY (2003) A comparative assessment of classification methods. *Decis Support Syst* 35(4):441–454. doi:10.1016/S0167-9236(02)00110-0
- Landis J, Koch G (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174. doi:10.2307/2529310
- Liu C, Frazier P, Kumar L (2007) Comparative assessment of the measures of thematic classification accuracy. *Remote Sens Environ* 107(4):606–616. doi:10.1016/j.rse.2006.10.010
- Muñoz JA, Martínez A, Verges J (1986) Thrust sequences in the Spanish Eastern Pyrenees. *J Struct Geol* 8(3–4):399–405. doi:10.1016/0191-8141(86)90058-1
- Ohlmacher G, Davis J (2003) Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Eng Geol* 69(3–4):331–343. doi:10.1016/S0013-7952(03)00069-3
- Poli S, Sterlacchini S (2007) Landslide representation strategies in susceptibility studies using weights-of-evidence modeling technique. *Nat Resour Res* 16(2):121–134. doi:10.1007/s11053-007-9043-8
- Powell RL, Matzke N, de Souza Jr C, Clark M, Numata I, Hess LL, Roberts DA (2004) Sources of error accuracy assessment of thematic land-cover maps in the Brazilian Amazon. *Remote Sens Environ* 90(2):221–234. doi:10.1016/j.rse.2003.12.007
- SAFELAND (2010) *Living with landslide risk in Europe: assessment, effects of global change, and risk management strategies*. Grant Agreement No.: 226479. <http://esdac.jrc.ec.europa.eu/projects/safeland>
- Santacana N, Baeza C, Corominas J, De Paz A, Marturià J (2003) A GIS-based multivariate statistical analysis for shallow landslide susceptibility mapping in La Pobla de Lillet area (Eastern Pyrenees, Spain). *Nat Hazards* 30(3):281–295. doi:10.1023/B:NHAZ.0000007169.28860.80
- Smits PC, Dellepiane SG, Schowengerdt RA (1999) Quality assessment of image classification algorithms for land-cover mapping: a review and proposal for a cost-based approach. *Int J Remote Sens* 20:1461–1486
- Stehman SV, Czaplewski RL (1998) Design and analysis of thematic map accuracy assessment: fundamental principles. *Remote Sens Environ* 64:331–344
- Sterlacchini S, Blahut J, Ballabio C, Masetti M, Sorichetta A (2008) A methodological approach for comparing predictive maps derived from statistic-probabilistic methods. *European Geosciences Union General Assembly (EGU), Vienna, Austria*. http://w3.unicaen.fr/mountainrisks/spip/IMG/pdf/poster_EGU_2008_Sterla.pdf. Accessed 9 Sept 2016
- Story M, Congalton RG (1986) Accuracy assessment: a user's perspective. *Photogramm Eng Remote Sens* 52(3):397–399
- Thiery Y, Malet J-P, Sterlacchini S, Puissant A, Maquaire O (2007) Landslide susceptibility assessment by bivariate methods at large scales: application to a complex mountainous environment. *Geomorphology* 92(1–2):38–59. doi:10.1016/j.geomorph.2007.02.020
- Van Den Eeckhaut M, Vanwalleghem T, Poesen J, Govers G, Verstraeten G, Vandekerckhove L (2006) Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). *Geomorphology* 76(3–4):392–410. doi:10.1016/j.geomorph.2005.12.003