

Document downloaded from:

<http://hdl.handle.net/10251/113372>

This paper must be cited as:

Cantarino-Martí, I.; Carrión Carmona, MÀ.; Goerlich-Gisbert, F.; Martínez Ibáñez, V. (2018). A ROC analysis-based classification method for landslide susceptibility maps. *Landslides*. 1-18. doi:10.1007/s10346-018-1063-4



The final publication is available at

<https://doi.org/10.1007/s10346-018-1063-4>

Copyright Springer-Verlag

Additional Information

A ROC analysis-based classification method for landslide susceptibility maps

Isidro Cantarino¹; Miguel Angel Carrion²; Francisco Goerlich³; Victor Martinez⁴

^{1,2,4} Department of Ground Engineering, Universitat Politècnica de València, Camino de Vera, s/n, 46071, Valencia, Spain

Email: vicmarib@trr.upv.es

³ Department of Economic Analysis, Universitat de València, Instituto Valenciano de Investigaciones Económicas (Ivie), Ivie, Calle Guardia Civil 22 Esc. 2 1º, 46020, Valencia, Spain

Abstract

A landslide susceptibility map is a crucial tool for land-use spatial planning and management in mountainous areas. An essential issue in such maps is the determination of risk thresholds. To this end, the map is zoned into a limited number of classes. Adopting one classification system or another will not only affect the map's readability and final appearance, but most importantly, it may affect the decision-making tasks required for effective land management. The present study compares and evaluates the reliability of some of the most commonly used classification methods, applied to a susceptibility map produced for the area of La Marina (Alicante, Spain). A new classification method based on ROC analysis is proposed, which extracts all the useful information from the initial dataset (terrain characteristics and landslide inventory) and includes, for the first time, the concept of misclassification costs. This process yields a more objective differentiation of risk levels that relies less on the intrinsic structure of the terrain characteristics. The results reveal a considerable difference between the classification methods used to define the most hazardous zones (in over 20% of the surface) and highlight the need to establish a standard method for producing classified susceptibility maps. The method proposed in the study is particularly notable for its consistency, stability and homogeneity, and may mark the starting point for consensus on a generalisable classification method.

Keywords: Landslide susceptibility maps, GIS, ROC analysis, classification systems

Acknowledgements

Authors acknowledge support from Ivie (www.ivie.es) research program, and Professor Francisco Goerlich funding from project ECO2015-70632-R

1. Introduction

Landslide susceptibility maps (LSMs) are powerful and essential tools for land-use planning and are especially suitable for mountainous environments. These maps are produced from predictive models based on complex and sophisticated mathematical methods (e.g. multi-criteria, discriminant, logistic regression, neural networks). They draw from comprehensive databases of the variables (also known as factors) involved in these landslide processes, notable among which are digital elevation models (DEM). The main reason for showing susceptibility data in map form is to produce spatial patterns that expose the underlying distribution of the data as accurately as possible (Fotheringham et al. 2000). An ideal classification system seeks a balance between the underlying information and the simplification of the continuous susceptibility values by dividing them into classes that will uncover intrinsic spatial patterns. To classify susceptibility, various methods are available that can simplify the information and aid interpretation of the map. These methods divide the histogram of LSM frequencies into a series of classes, the number of which is usually set by the user.

The main aim of classification systems is to create cartographic representations with optimal visualisation and to simplify their interpretation. However, aggregating data in a given number of classes can have an adverse effect on the apparent result, depending on which criteria are used to create the map. For example, simply changing the boundaries between classes can produce maps with very different appearances (Evans 1977). The overall accuracy of the classified map therefore determines the reliability of the data for any application (Liu et al. 2007). The consequences of the classification must be carefully analysed because of the crucial role it can play in decision-making tasks (Kiang 2003). Meaningful, consistent measures of map accuracy are needed to evaluate how suitable a map is for application in a particular case.

However, scant research evaluating landslide susceptibility offers information on the various classification methods available. It is therefore important to adopt measures that analyse the suitability of a given classification strategy (Powell et al. 2004) and to compare two or more discretisation techniques (Foody 2004). Neither does any one method stand out as most suitable for general use. Indeed, in their comprehensive review of the landslide mapping and GIS literature, Chacón et al. (2006) conclude “*there is a lack of internationally accepted classifications and a common standard would be highly valuable for comparing maps, and also to classify the landslide areas all around the world*”.

Many of the approaches used in the literature to test the accuracy of landslide spatial prediction maps (Stehman and Czaplewski 1998, Smits et al. 1999) can also be applied to assess the suitability of classifications in susceptibility maps. The ‘error matrix’ (also known as the confusion matrix or the contingency table) is key to most measures of thematic map accuracy (Powell et al. 2004; Gupta et al. 2008). Once the mathematical model has adopted a prediction threshold, the binary prediction (failed/correct) can be compared with the observed landslide. The error matrix can then be constructed showing the number of correctly and incorrectly predicted observations. Receiver operating characteristic (ROC) analysis is also based on the sequential study of the error matrix, a subject dealt with later in the article.

Landslide susceptibility maps generally classify areas into various levels of potential landslide occurrence, from very high to very low levels of susceptibility, represented on an ordinal scale. Although some maps have been published with eight levels (Nadim et al. 2006), they usually have between two and five classes. In this study, a distribution of five classes was applied to analyse the correspondence of the susceptibility levels obtained from different classification systems.

Susceptibility data can be categorised either with common classification schemes available in commercial software packages or created manually. However, when the susceptibility values calculated with one of these procedures are transferred to a map with classes established by any system, the maps that originally had the same predictive power no longer have the same meaning. To elucidate this issue, therefore, in this study the same physiographic dataset was used to produce susceptibility maps via two different methods and, at the same time, apply different classification systems. As well as the required error or confusion matrix, a series of quantitative indices was calculated for the comparative analysis in order to identify any significant differences between the conventional classification systems and the one proposed in the present article, and to determine which one yields the best performance.

2. General methodology

2.1. Introduction

The various approaches taken to prepare LSMs use classification systems to group the data in order to construct the final susceptibility map. Research has centred on the predictive ability of these classification systems, which is reflected in the models' reliability and accuracy. However, no studies have examined which classification system is the most suitable for defining the levels of susceptibility to be included in the map. Classification implies a loss of information or a different redistribution of the data, depending on which method is applied.

It is important to determine how the classification affects the distribution of susceptibility and its reliability as compared to other possible classification methods in the area studied. Chung and Fabbri (2003) contend that in LSMs represented by continuous value pixels, the division between classes is no trivial matter and must be treated as one of the most important stages in modelling susceptibility.

There are currently no standard rules to classify landslide susceptibility maps (e.g., number and name of classes, class boundaries). One of the methods authors sometimes adopt to divide the susceptibility histogram into categories is based on expert criteria (Ohlmacher and Davis 2003, Van Den Eeckhaut et al. 2012) and does not take into account the real underlying data. This data categorisation cannot be automated or statistically tested (Ayalew and Yamagishi 2005). Therefore, it could be said that all the methods commonly used to classify landslide susceptibility maps cannot be considered as a definitive, generalisable solution.

Today, cartographic representation in regions with large amounts of data requires an automatic, objective process. Many of the classification methods can easily be applied by means of geographic information systems (GIS), and some of them are integrated into the GIS software (Zhu 2016). In the present study, some of these methods have been applied to zone different LSMs and draw comparisons. The resulting classes group the values of the landslide susceptibility index (LSI) calculated for each pixel according to the specific criteria for each method.

In this paper, landslide data was used to classify the susceptibility maps into five susceptibility classes commonly used in small-scale landslide susceptibility zonation (e.g., Guzzetti et al. 2006; Van Den Eeckhaut et al. 2012): very low (VL), low (L), moderate (M), high (H), and very high (VH). These categories are considered sufficient to reveal any spatial patterns existing in a dataset and aid interpretation of these LSMs (Armstrong et al. 2003). This approach has often been used to evaluate the concordance of binary maps; VH and H are generally accepted to define potentially unstable zones, while L and VL are regarded as stable.

In the following section, the most common classification systems for susceptibility maps are reviewed. The bases for a new classification system based on ROC curve analysis are then proposed. Finally, this new classification and the other

classifications analysed are applied in a mountainous area in the northeast of the province of Alicante (Spain), and their suitability and reliability are tested.

2.2. General classification systems

The classification systems commonly applied in the literature on susceptibility are either calculated automatically by GIS software, or defined by the user or based on expert criteria. The systems replicated in this article all belong to the former type and are described below.

1. Natural Breaks (NB).

This is an automatic method usually integrated into GIS software, such as ArcGis by ESRI or QGis (free open source software), possibly the most widely used GIS. The classes are based on natural groups inherent in the data and the boundaries are established statistically when relatively large jumps appear in the values of the susceptibility data determined by their variance. It makes a classification by minimising the variance value within one class and increasing the value of the variance between classes. Based on natural breaks, Jenks (1967) developed an optimisation method to minimise within-class variance while maximising between-class variance.

2. Quantiles (Q).

This system is also integrated in GIS software and designates the area of equal coverage, assigning the same number of cells to each class, and therefore occupying equal areas. In this case, the range of possible susceptibility values is divided into intervals of unequal sizes. This type of classification adapts well to linearly distributed data and allows the average values at the extremes to be highlighted in wider intervals. Its main drawback is that very different values with no clear grouping may be located in the same class.

3. Head/Tail Breaks (HT).

This classification method, devised by Jiang (2013), can be used to establish groupings or hierarchies for data with a heavy-tailed distribution. Heavy-tailed distributions are strongly right skewed with a marked asymmetry between large and small values, usually characterised by a power law, a lognormal or an exponential function. These distributions are often found in many social or geographical phenomena. As Jiang (2013) describes, "*The head/tail method groups the data values into two parts around the arithmetic mean and continues the partitioning for values above the mean iteratively until the head part values are no longer heavy-tailed distributed*". Thus, in a digital elevation model (DEM) low elevations far outnumber high ones; this would be a typical heavy tailed distribution. A similar case is the distribution of susceptibility values, to which it seems to adapt satisfactorily. However, in a study of susceptibility comparing NB with HT, Basofi et al. (2015) found that the low susceptibility values includes too many data with the HT classification, and is less significant than NB. The present study therefore attempts to provide a better adaptation of the HT method for the five classes described above. This adaptation entails splitting the head level; that is, starting from the arithmetic mean the three tail levels are calculated and two are added at the head (2H/3T) rather than the four proposed in the original method (1H/4T). This classification, referred to here as the modified HT (HTm), is applied in this article.

Many other methods of classification have also been used in risk classification studies such as equal interval and, in particular, standard deviation. The latter shows the degree to which pixel values deviate from the mean: these values are then used to create class breaks; as it is based on variance it can present similarities to NB. Indeed, the asymmetry of the present dataset and its tail breaks lead to similar results to this latter classification, and as a result, it has not been considered in this study.

2.3. Classification method based on ROC analysis

Receiver operating characteristic curve analysis, or simply ROC analysis, provides tools to differentiate two classes, established through a diagnostic test, in an optimal manner. This analysis is based on the final distribution of a classification method that differentiates between correct and failed predictions according to a 2x2 contingency table:

Table 1. Contingency table

		True condition (observed)	
		Positive (unstable)	Negative (stable)
Predicted condition	Positive (unstable)	True positive (TP) Correct	False positive (FP) Failures or false alarms Commission or Type I error
	Negative (stable)	False negative (FN) Omission or Type II error	True negative (TN) Absences

The ROC curve plots the true positive rate (TPR), indicating that the classifier locates landslides successfully, and the false positive rate (FPR), showing landslides located where they do not exist, for a set of threshold or cut-off values. TPR is the *sensitivity* (Se), and 1-FPR is the *specificity* (Sp). The specificity is the true negative rate (TNR) which correctly locates stable zones or those with no landslides. The accuracy of this diagnosis depends on how well the test separates the group being tested into elements with and without the feature in question. Accuracy is generally measured by the area under the ROC curve (AUC). This area can be interpreted as the probability that a classifier gives a randomly chosen positive event a higher score than a negative event.

The classification proposed in this article is based on finding the optimal cut-off point(s) (c1, c2, ..) that best group the available data, starting from the test or inventory data. The first stage aims to differentiate two main classes: potentially unstable and stable terrains. One of the most common ways of obtaining this dichotomous differentiation in ROC analysis is with the Youden index (Y) or *informedness*, the expression for which is:

$$Y = TPR - FPR = \text{Sensitivity (TPR)} + \text{Specificity (TNR)} - 1 \quad [1]$$

Mathematically, Y reflects the intention of maximising overall correct classification rates and thus minimising misclassification rates to find the optimal cut-off point; that is, the point that maximises the difference between TPR and FPR, or where sensitivity and specificity are at their highest, marks the optimal cut-off point c, $Y(c) = \max \{Se + Sp\}$. The index gives equal weight to false positive and false negative values, and can be generalised by introducing different weights (perhaps based on the cost of the different classification error types) for these types of errors.

This two-group division in a susceptibility map is an oversimplification and does not fully exploit all the information from the ROC analysis (Greiner 2000). First, these maps require a limited set of differentiated susceptibility levels

(usually four or five); and second, dividing terrain into classes of susceptibility has economic and social consequences or “costs”, that must be taken into account. For example, areas classified as stable have a higher economic value than those classified as unstable. However, with the exception of Frattini (2010) for a dichotomous stable/unstable classification in a mountainous area in the north of Italy, none of the techniques used in the literature to assess the accuracy of landslide susceptibility models take account of misclassification costs. This limitation is significant for landslide susceptibility analysis as the costs of misclassifications differ greatly depending on the error type.

More generally, Zweig and Campbell (1993) propose a method to calculate the optimal point by maximising the correct classifications and considering the costs of misclassifications and correct classifications. Taking into account only the cost of the false positive errors (C_{FP}) and omissions or false negatives (C_{FN}), the optimal cut-off value c in the ROC curve will be the one that maximises the function:

$$\max \left\{ Se - \frac{C_{FP}}{C_{FN}} \times \frac{1-p}{p} \times (1 - Sp) \right\} \quad [2]$$

where p is the prevalence or presence of landslides over the whole area studied (pre-test or a priori probability). When equal costs and $p = 0.5$ in Equation [2] are considered, it coincides with the Youden index calculated in [1]. Function [2] is known as the generalised Youden or GY (Geisser 1998; Greiner et al. 2000; López-Ratón 2014). This index is recommended when the weights or costs of Se and Sp are presumed to be different, as in the case of landslide phenomena. Thus, an optimal cut-off c is the one that maximises expression [3], for which prevalence and costs must be considered in order to calculate GY properly, that is,

$$GY(c) = Se(c) + \frac{1-p}{p} \times \frac{C_{FP}}{C_{FN}} \times Sp(c) - 1 \quad [3]$$

Another criterion has been proposed for selecting the optimal cut-off point, based on minimisation of a term that measures the cost of incorrect classifications. This is the misclassification-cost term (MCT), also known as the minimum loss function criterion or expected cost (Metz 1978; Drummond 2006).

$$MCT(c) = \frac{C_{FN}}{C_{FP}} \times p \times [1 - Se(c)] + (1 - p) \times [1 - Sp(c)] \quad [4]$$

Because inventories are incomplete, and LSMs include cells where landslides could never occur (due to slope, soil use, and so on), the pretest probability will be low ($p = 0.1$ or less). It is not advisable to establish optimal cut-off points with these low prevalence values since, although many correct positive TP are identified in high susceptibility classes (with few false positives), many potentially unstable cells are also left in low susceptibility classes. As these false negatives are located in zones with greater human activity, they are particularly important in terms of the potential public concern they might generate.

In sum, it is difficult to determine the real prevalence of the initial population, and the prevalence of the sample is insufficient as it is over-optimistic. Unfortunately, there is a large margin of uncertainty in the actual distribution of landslides used a priori to classify units into positives or negatives. Greiner et al. (2000) solve this uncertainty by noting that “... generally, without better information, one tends to assume $p = 0.5$ ”. Likewise, Frattini et al. (2010) adopt a condition of equal probability in their application in the north of Italy. Working with this value, only a different

comparative study of costs or importance is needed for each section or class in order to calculate the optimal cut-off points c_1, c_2, \dots (López Ratón et al. 2014) according to function [5].

$$GY(c_1, c_2, \dots) = Se(c_1, c_2, \dots) + \frac{C_{FP}}{C_{FN}}(1, 2, \dots) \times Sp(c_1, c_2, \dots) - 1 \quad [5]$$

Therefore, the relative importance (or 'cost') of failed classifications or false alarms (FP) and omissions (FN) must be differentiated for each class. By establishing different misclassification costs these optimal cut-off points can be defined as the boundaries of each class.

The likelihood ratio (LR) is a useful way to achieve an approximate separation and characterisation of classes. This ratio, expressed as a percentage, quantifies the improvement in a classification test as the increase in the post-test probability of correct classifications in the intervals classified on the initial pre-test probability. Specifically, working with correct presence (+) or absence (-) classifications gives expressions [6]

$$LR(+) = \frac{TPR}{FPR} = \frac{Se}{(1 - Sp)} > 1 \quad LR(-) = \frac{FNR}{TNR} = \frac{(1 - Se)}{Sp} < 1 \quad [6]$$

To quantify the improvement in the classification of an LSM in general terms, an illustrative example of an LSM with ROC analysis is proposed. A 10% probability of landslides is assumed across the whole area with an area under curve (AUC) of 0.8, dividing the ROC curve into three sections. In this example, according to Bayes' theorem and the tables devised by McGee (2002), for the sections with high LSI values, the probability of a correct classification in unstable pixels would be higher than 50% (LR+), whereas in the low sections it would fall below 5% (LR-), that is, a 95% correct classification for stable zones.

Table 2 displays the characteristics of the main zones and a rough estimation of relative costs based on implementing stabilisation elements for slopes and damage to vulnerable elements.

Table 2. Characteristics of the main classes

Zone	C_{FP}/C_{FN}	Comments
High VH – H	> 1	Potentially unstable section: fairly efficient for defining zones with possible landslides. High C_{FP} due to unnecessary slope stabilisation. Low C_{FN} due to limited presence of major infrastructure or dwellings.
Moderate M	~ 1	Intermediate zone that defines neither stable nor unstable zones adequately. Similar failure and omissions costs.

Low	Stable section: efficient for ruling out zones with landslides. High C_{FN} as it is in a safe zone with main infrastructures and especially dwellings: zero public tolerance for these errors. Low C_{FP} due to no implementation of stabilisation elements.
L – VL	< 1

It is important to note that the application of ROC curves to landslide models should be approached with some caution. Broadly, the percentage of false positives is inversely proportional to a model's efficiency (a model that raises false alarms of a hurricane, for example, or other catastrophic events). This is not the case for landslides. A high percentage of false positives does not necessarily constitute an adverse condition. According to Carrara et al. (2008), "false positives may indicate areas where landslides were not detected because of concealment by farming activity or subsequent erosion or vegetation cover. It is well known that landslide inventories are incomplete, error prone documents".

Because inventories are incomplete and inaccurate, false negatives (FN) are more relevant in LSMs than false positives (FP), an issue that the Youden index obviously does not take into account. In their study, Frattini et al. (2010) divide the misclassification cost for false positives (FP) by a corrective factor of 2 to offset the uncertainty in the a priori classification of stable terrain units, under the assumption that half of the terrain units may be misclassified.

In sum, to apply equation [5] a set of costs must be established to determine the optimal cut-off points that limit and define each susceptibility class. First, in line with the above, the real cost of failures (C_{FP}) must be adjusted (C_{FPaj}) by the reduction coefficient k_r :

$$C_{FPaj} = k_r \times C_{FP} \quad [7]$$

The ratio C_{FPaj}/C_{FN} should also be maintained as in Table 2.

It seems reasonable that there will not be so many failures (FP) in the low susceptibility zones as to require the application of the reduction coefficient equal to 0.5 proposed by Frattini et al. (2010). Taking the five levels defined above, no adjustment is made for class L ($k_r=1$), and for class M it is $k_r = 0.75$. For the remaining levels (H and VH) $k_r = 0.5$ is applied as these are high susceptibility zones where the highest number of erroneous false positives are expected. The costs of omission errors (FN) are not corrected since, according to Frattini et al. (2010), zones interpreted as unstable present strong evidence and are usually well classified.

In line with the above and the characteristics of these classes shown in Table 2, the lower boundaries of each class can be defined by estimating the evolution of their C_{FP} and C_{FN} costs as shown in Fig. 1. These costs experience rapid variation in the extremes that adjust better to an exponential than to a linear or quadratic function. Thus, a geometric progression of costs according to a type $C_F = 2^{f(LSI)}$ exponential equation was assumed in this study. Another simplification is that the cost curves have a symmetric distribution with an intersection in the middle zone. The cost function of FP rises, taking into account the gradually increasing complexity of performing and accessing stabilisation work as the LSI increases (and with it, the slope gradient). The cost function of FN falls, because the value of the land, infrastructures, buildings, dwellings etc. rapidly decreases, as does the population. These cost curves are developed in the interval [0,8] since for the values (1, 2, 4, 8) were assigned to the four main points, following the geometric progression indicated, and are displayed in Fig. 1. Obviously, this proposal is formulated for a generalist hypothetical application as its specific calculation depends on each study area, and a complex analysis would therefore be required.

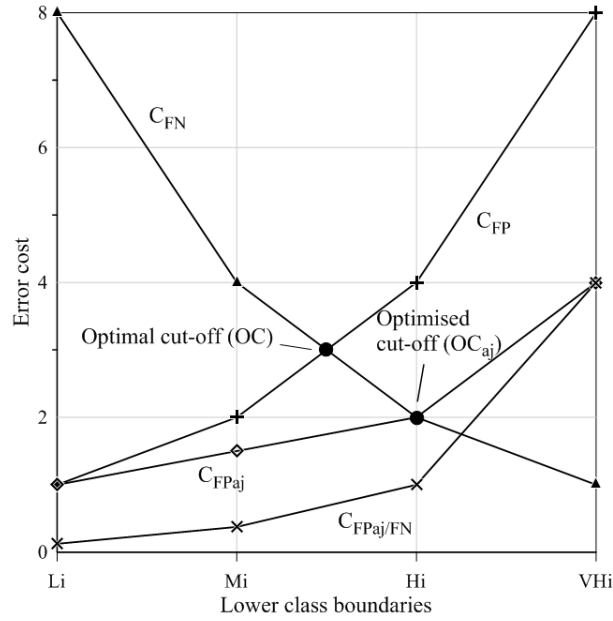


Figure 1. Cost curves estimated indicating the lower class boundaries

The point of intersection between C_{FP} and C_{FN} curves in Fig. 1 denotes equality $C_{FP} = C_{FN}$ and, therefore, the Youden index expressed in [1]. As defined above, it is an optimal cut-off (OC) point to separate unstable and stable zones, and it should be found in the middle zone of class M. However, the FP adjustment (C_{FPaj}) means it shifts towards the boundary between classes M and H (Hi), more in line with Table 2. By substituting the C_{FPaj}/C_{FN} costs relation from Fig. 1 in function [5], the values of the lower boundaries can be calculated for the five levels as shown in equation [8]

$$GY(c) = \max \left\{ Se(c) + \frac{C_{FP}}{C_{FN}} \times Sp(c) \right\} \quad [8]$$

Equation [8] expresses a binary classifier based on maximising success rates and attending to misclassification costs. A similar approach is exposed in [5] with the MCT criterion; this classifier minimises the error rates FPR and FNR (false negative rate), and its expression [9] yields the same results as [8].

$$GY(c) = \min\{MCT(c)\} = \min \left\{ FNR(c) + \frac{C_{FP}}{C_{FN}} \times FPR(c) \right\} \quad [9]$$

When equation [8] is applied, the resulting values maximise the functions of each class and define their lower boundaries. The functions of each level are presented in Table 3.

Table 3. Lower class boundaries according to GY

Class	C_{FPaj}/C_{FN}	Function (1)
Very High (VH)	4	$\max \{Se + 4 \times Sp\}$ (2)

High (H)	1	$\max \{Se + Sp\}$ (3)
Moderate (M)	3/8	$\max \{Se + 3/8 \times Sp\}$
Low (L)	1/8	$\max \{Se + 1/8 \times Sp\}$
Very Low (VL)	~ 0	$\min \{LSI\}$

(1) LSI value that maximises the function

(2) The upper value obviously is $\max \{LSI\}$

(3) Equals to Youden index

Figure 2 shows the situation of these cut-offs calculated on a theoretical ROC curve, including the plot of the cost functions from Table 3, normalised within the interval [0,1] for representation.

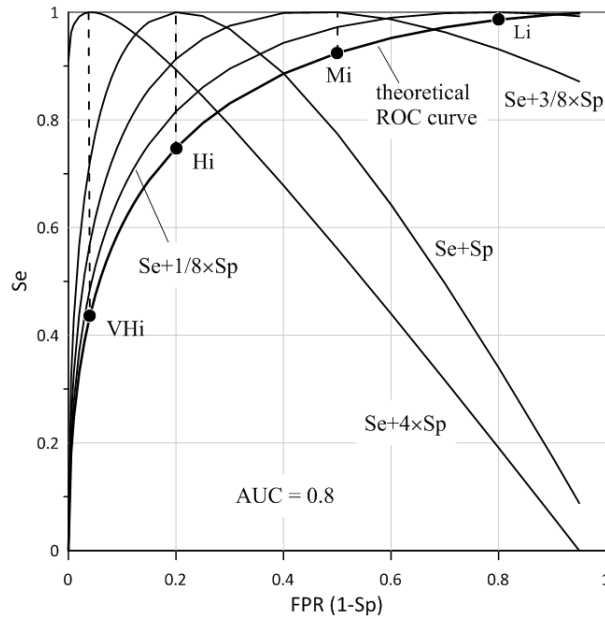


Figure 2. Theoretical ROC curve with class boundaries

Clearly, this method for calculating susceptibility boundaries in an LSM allows us to make a selection of the cells that present a certain instability threshold without substantially altering the results. As the horizontal projection of the ROC curve advances, the increase in the AUC index value slows down until the saturation point is reached and it no longer increases. This point should define minimum lower level of any LSM, as further reductions in the level do not improve the adjustment.

3. Case-study: Landslide Susceptibility Map (LSM) of La Marina

3.1. Description and data sources

The area selected for the comparative study is located in La Marina (Fig. 3), a district in the province of Alicante (southeast Spain) by the Mediterranean Sea. The area includes 51 municipalities and covers 1,339 km²; it has a population of 363,630 inhabitants according to the 2011 census (Spanish National Institute of Statistics, INE). It is a populated mountainous environment ranging from sea level to around 1,500m. Its relief is shaped by its proximity to the sea, with a river system that deeply dissects the territory.

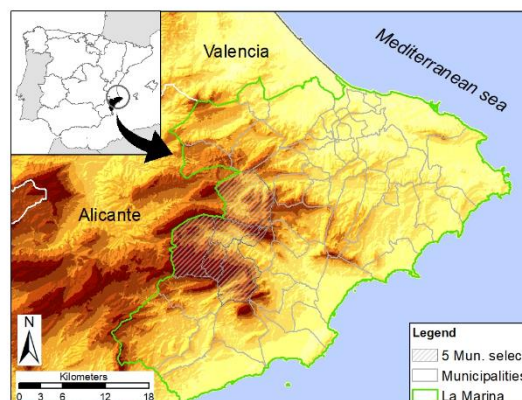


Figure 3. Study area map

The province of Alicante, where La Marina is located, is the area of the Valencian Community with the highest landslide rate per surface unit (Hervás 2017). Its mountainous terrain means the coastal strip is not exempt from risk, a situation that is heightened by its high tourism value and residential occupation. Five of the 51 municipalities in the region are located in the most mountainous inland area, which covers 151.5 km² (see “5 Mun. select” in Fig. 3). According to the inventory, this is the zone with the highest density of landslides per surface unit.

The lithology of the mountainous zones in the study area is essentially carbonate (limestone and dolomites with substantial banks of marl) from the Late Cretaceous period; the foothill zones consist of later Tertiary and Quaternary clay and silt material. The incidence of intense Alpine tectonics causes a high level of rock fracturing, favouring slope instabilities.

The data for the present article were those used to define the susceptibility factors and those representing the terrain characteristics. All data came from the official national sources detailed below.

1. Information on terrain slope gradient was derived from a digital elevation model (DEM) MDT25 downloaded from the Spanish National Geographic Institute (IGN). The database was produced with a Lidar flight and has 25x25m pixel size and a resolution of 50 cm/pixel.
2. Land Use and Land Cover (LULC) was obtained from the IGN's SIOSE (Spanish Land Use and Land Cover Information System) 2006 model (1:25,000 scale). Because this model provides very thorough and complex information coverage, a hierarchical model derived from the SIOSE was selected, which is more straightforward to apply (Cantarino 2010).
3. Lithology. The information on the surface geology was taken from the 1:50,000 scale single legend National Geological Map (MAGNA) produced by the Geological and Mining Institute of Spain (IGME).

The susceptibility maps were compared and validated using inventory data from an official regional organisation and the European landslide map.

1. Digital landslide inventory data (Landslide databases, LDBs) were from a 1:50,000 scale vector format landslide map produced by the Department of Public Works of the Valencian Regional Government (COPUT 1998), using 1:50,000 scale geological and geotechnical maps from the Spanish Geomining Technical Institute (ITGME), and aerial photographs available at that time. Data from specific publications and field surveys were also used to observe landslide events through morphology, lithology, structure, slopes, vegetation, etc.
2. European Landslide Susceptibility Map (ELSUS 1000) v1. This map shows levels of spatial probability of generic landslide occurrence at the European scale. ELSUS was produced by the Joint Research Centre (JRC) in 2012 with a 1x1km grid and five susceptibility classes (VH, H, M, L, VL) for various types of physiographical regions (coasts, plains and mountains). Zone five, classified as mountainous with an arid climate, was selected to compare with the area studied in this research (Gunther et al. 2014)

3.2. Procedure design

The procedure followed to compare the classification systems in the selected example is reported in Fig. 4. Calculating susceptibility maps requires, first, the calculation of the weights of each factor involved (frequency ratio, FR). The LSI value for each pixel was obtained by two models: spatial multi-criteria evaluation (SMCE) and logit (see sections 3.4.1 and 3.4.2). These LSI values were used to generate the LSMs, which were then validated by ROC analysis and zoned using four classification methods (GY, NB, Q and HTm; see sections 2.1 and 2.2). These classified maps were then analysed using four statistics that define four main qualities, namely, the stability ratio (Rst), the landslide density index (Rind), the Kappa index, and Moran's index of spatial autocorrelation (Moran). These indices allow subsequent grouping through cluster analysis. Each group or cluster presents certain specific characteristics that lead to the final conclusions.

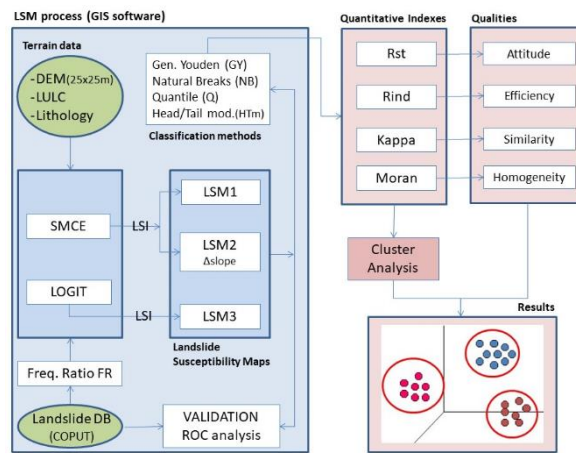


Figure 4. Workflow chart (explanation in text)

3.3. Assignment of factor class weights

Before preparing the susceptibility maps, each factor (variable) must be individually analysed by means of a bivariate statistical approach. This allows us to understand the real influence of each variable on the physical phenomenon and to classify the variable map accurately, yielding a direct or inverse proportionality with landslide occurrence. This preliminary step is especially necessary in case of non-numerical variables such as geolithological features and the land-use map. A bivariate statistical approach shows the landslide occurrence is dependent on a single variable and the conditional probability is obtained via Bayes' theorem:

$$p(\text{landslide}/X) = \frac{P(\text{landslide in } X)}{P(X)} \quad [10]$$

This conditional probability is equivalent to factor class weights, which in this study were allocated directly through the landslide frequency ratio (FR). These approaches are based on the relationships observed between landslide distribution and each landslide-related factor, revealing the correlation between landslide locations and the factors in the study area. The frequency ratio (FR) is the ratio of the area affected by landslides to the total study area, and is also the ratio of the probabilities of landslide occurrence to non-occurrence for a given attribute.

FR is calculated from the analysis of the relation between landslides defined in the COPUT landslide inventory and the attributed factors. Thus, for each class j of factor i :

$$FR_{ij} = \frac{A_{ij}/B}{C_{ij}/D} \quad [11]$$

where:

A_{ij}: number pixels with landslide for each factor i and class j

B: total number pixels with landslides in the study area

C_{ij}: number of pixels in class area j of factor i

D: total number of pixels in the study area

The FR value may be higher than 1, and further indicates that the ith class of factor F (F_i) favours landslide occurrence. Based on the inventory data, FR values were obtained according to [11], then normalised in the 0-1 range and multiplied by 1000 (FR_n) for the discrete variables (lithology and land use).

The normalised FR values were then modified by applying expert knowledge to account for over- or underestimated weight values based on parameter class sizes and landslide distributions to obtain the results reported in Table 4. The modifications introduced were limited so that their variations in absolute and relative value were not too important. These values are given in column FR_{adj}.

Table 4. FR values for non-continuous factors (discrete variables)

Class	Lithology	%	FR _n	FR _{adj}	LULC	%	FR _n	FR _{adj}
1	Massive hard rocks	4.1	155	135	Urban	8.3	25	15
2	Alternations of hard rocks	11.4	97	124	Artificial non-residential	4.0	59	35
3	Non-hard rocks	0.7	364	315	Artificial industrial	0.7	71	40
4	Marls and clays with limestone	39.5	152	200	Arable crops	1.2	16	10
5	Soft rocks: marls, chalks	20.5	64	96	Irrigated orchards	17.3	19	28
6	Loose Tertiary materials	0.5	148	100	Non-irrigated crops	4.2	49	72
7	Loose Quaternary materials	23.3	20	30	Forest	12.1	171	210
8	--	--	--	--	Scrubland	20.6	366	320
9	--	--	--	--	Grasslands, bare ground	31.6	223	270
Total			1000	1000			1000	1000

Slope is a continuous variable; here classes can be established more automatically and the definition of classes in this factor is less subjective. Langping et al. (2017) suggest calculating the frequency ratios for every identical normalised factor value rather than every factor class (conventional FR method). In the present study, FR is calculated for a large number of regular classes, which are then grouped by irregular intervals according to ranges of variation of the FR value, and the FR calculation is repeated for these new classes. This procedure differs from the method proposed by Langping, which requires a constant bin width. The method proposed in this article, referred to as the Grouped Frequency Ratio (GFR), seeks to provide a better fit to the data structure and to the most relevant slope intervals, and to reduce the number of classes in a rational manner.

In the example given, the degree slope raster is reclassified in two degree intervals, yielding 41 classes. The FR are then calculated, normalised to 1 and multiplied by 1000 (FRn). The process continues with the grouping of FRn values that are close to each other in irregular intervals, selecting a bin width of 10 for slopes above 28°, and proportionally smaller for lower slopes. The FR are then recalculated for the eight defined classes, yielding a GFR value, finally normalised over a total of 1000 (GFRn). The distribution of classes by surface is therefore more homogenous and avoids a large development in the lower classes.

Table 5. GFRn values for the slope factor

Class	FRn	Bin width	Slope (°)	GFRn	Surface (%)
1	0.0 – 0.5	0.5	0 – 6	2	27.8
2	0.5 – 2.5	2.0	6 – 16	10	30.5
3	2.5 – 5.0	2.5	16 – 22	25	15.1
4	5 – 10	5	22 – 28	54	12.5
5	10 – 20	10	28 – 32	108	6.1
6	20 – 30	10	32 – 38	188	4.9
7	30 – 40	10	38 – 48	265	2.2
8	40 – 50	10	48 – 90	348	0.9
TOTAL				1000	100.0

According to the GFRn value in Table 5, the first class will be a zone with almost no landslides, since the probability of one occurring in this Class I is $P(C1) = \text{GFRn}/1000 = 0.002$. Although it is generally considered (according to Varnes 1984) that no landslides should occur below 4° (equivalent to 3.5%) in a mild dry climate such as that of La Marina, this class is a fairly adequate presentation of an area with no presence (or “safe area”), and accounts for more than a quarter of the surface studied (27.8%). Areas above 48° can be grouped into a single class because the landslide mechanism does not directly correspond to the hillside slope.

3.4 Preparation of susceptibility maps

A 25x25m pixel or cell was used as the surface unit for the study in all the maps. Two methods commonly applied in the susceptibility map literature were used to obtain the susceptibility values in each pixel: the spatial multi-criteria method (SMCE) and logit regression.

3.4.1 Spatial multi-criteria method (SMCE)

The first assessment follows an approach based on spatial multi-criteria evaluation (SMCE) discussed by Malet et al. (2013) for landslide susceptibility zoning that allows the integration of landslide information and expert knowledge. In the SMCE, first the factor weights attributed to landslide susceptibility were established in an analytic hierarchy process (AHP) through pairwise comparisons (Saaty 1980) of the three criteria: slope gradient, lithology and land cover. In the pairwise comparison, a value between 9 (far more important than) and 1 (equally important as) can be assigned to each pair of parameters in a comparison matrix by rating rows against columns. The weights computed via the pairwise comparisons can be evaluated by means of a consistency ratio (CR) that indicates the probability of the decision matrix being generated randomly.

The weighting used by ELSUS was first considered as a pattern for comparison. However, the data used to calculate the slope in ELSUS do not have the same resolution as the DEM from the IGN, and a higher weighting should therefore be assigned in this case; for this reason, the ELSUS values are not used in this article. LSM1 and LSM2 were produced for this study with AHP, LSM1 being the option obtained by the expert panel, while in LSM2 the slope was strengthened at the expense of the lithology, the factor that presented the lowest resolution (1:50.000), followed by land use (1:10.000). Previous studies (Van Den Eeckhaut et al. 2012; Günther et al. 2013; Jaedicke et al. 2014) have demonstrated that slope gradient is *“by far the most important parameter to determine general landslide susceptibility at the continental scale”* (Günther et al. 2014). It is useful to have a range of values for the same calculation method in order to determine its sensitivity through an analysis of differences in the results.

Table 6 displays the weights assigned to the factors according to Saaty’s (1980) AHP method. The consistency ratio of comparison matrices (CR) was less than 0.10 for all cases, which is acceptable for the AHP analysis to proceed.

Table 6. Factor weight assignment based on AHP

Variables	Weights		
	ELSUS	LSM1	LSM2
Slope	0.58	0.65	0.74
Lithology	0.28	0.23	0.16
LULC	0.14	0.12	0.10

These coefficients for each factor and the weights obtained for each class according to GFR/FR enable us to obtain the LSI i values making up each susceptibility map LSM1 and LSM2. These values were calculated through a weighted linear sum of the parameter and parameter class weights with equation

$$LSI_i = \sum_{j=1}^{n=3} w_j \times x_{ji} \quad [12]$$

where w_j is the weight of the parameter j (AHP) and x_{ji} is the weight of the parameter class i in criterion j (GFRn or FRn)

3.4.2. Logit regression

Logistic regression is a type of regression analysis used to predict the result of a categorical variable (a limited number of categories can be adopted); in the case of a susceptibility map binary logistic regression is applied, 1 or 0 according to the independent variables or predictors considered as terrain characteristics (slope, lithology and land use). It is also of interest to compare the resulting maps with those obtained by SMCE to analyse consistency.

The multivariate regression relation allowing the spatial relationship to be established between landslide occurrence and its conditioning factors is expressed by the following linear equation:

$$Z = b_0 + b_1 \times X_1 + \dots + b_n \times X_n \quad [13]$$

where Z is the dependent variable that expresses the presence (1) or absence (0) of landslides; b_0 is the intersection; $b_1 \dots b_n$ are the estimated coefficients that measure the contribution of the independent variables ($X_1 \dots X_n$) and define the factors that most affect stability (Duman et al. 2006).

The main advantage of logistic regression is that by adding a link function, all the variables can be continuous, discrete, or a combination of the two without necessarily having to have normal distributions (Lee et al. 2007). The log link function can be applied in cases where the dependent variable is dichotomous (Atkinson and Massari 1998). This yields the probability (P) that the dependent variable (Z) will take the value 1 defined in [14].

$$P = \frac{1}{(1 + e^{-Z})} \quad [14]$$

Before applying the logistic multiple regression, each variable must be analysed individually. When a logistic model is adopted, the logit value is calculated through a linear function of the independent variables X. The same factors (variables) were used as in SMCE, again assigning the FR values (GFRn and FRadj) previously calculated in section 3.3 for each one of them. The results of the analysis are presented in Table 7.

Table 7. Logit regression coefficients

Variables	Coef.	Std. Error	z	P> z	[95% Conf. Interval]	
LULC	0,00436	0,000037	117,93	0	0,00429	0,00443
Slope	0,01552	0,000034	455,62	0	0,01545	0,01559
Litho	0,00325	0,000050	65,20	0	0,00315	0,00334
Constant	-4,80017	0,010919	-439,64	0	-4,82157	-4,77877

Map LSM3 is obtained from the LSI values calculated by applying equation [15], according to the logit regression coefficients from Table 7.

$$LSI = P \times 1000 = \frac{1000}{1 + e^{-0.00436LULC - 0.01552Slope - 0.00325Litho + 4.80017}} \quad [15]$$

The probability values go up to 1000 and range between 10 and 953. Logically, they coincide with the minimum and maximum values of the factors used, and are radically different from those obtained with the SMCE methods.

3.5. ROC curves

ROC curves were generated for the three models with the LSI data calculated according to [12] (LSM1, LSM2) and [15] (LSM3) and data from the inventory, yielding AUC values and the class boundaries for the GY classification. As described in section 2.2, the AUC index is used to evaluate the quality of the model fit. This validation is essential in order to determine the predictive ability of the models used and to compare them. The criteria in section 2.2 and from Table 3 were applied to the ROC curves, from which the optimal cut-off point was obtained to define the class boundaries. Figure 5 displays all these data.

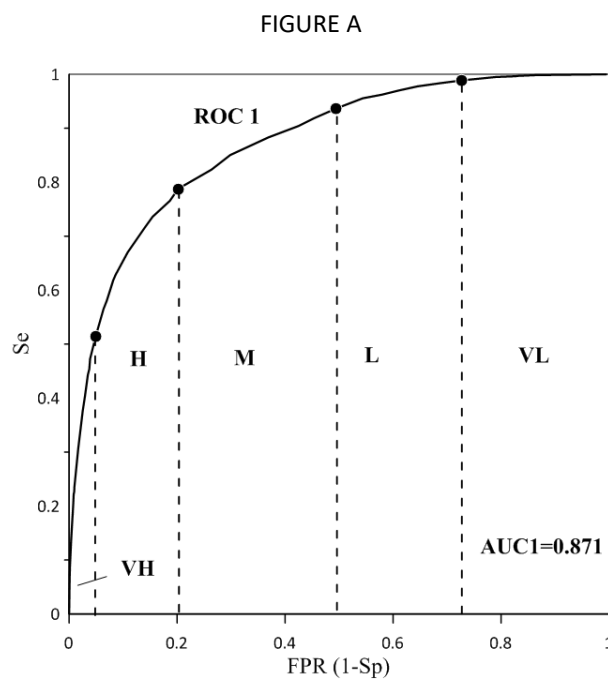
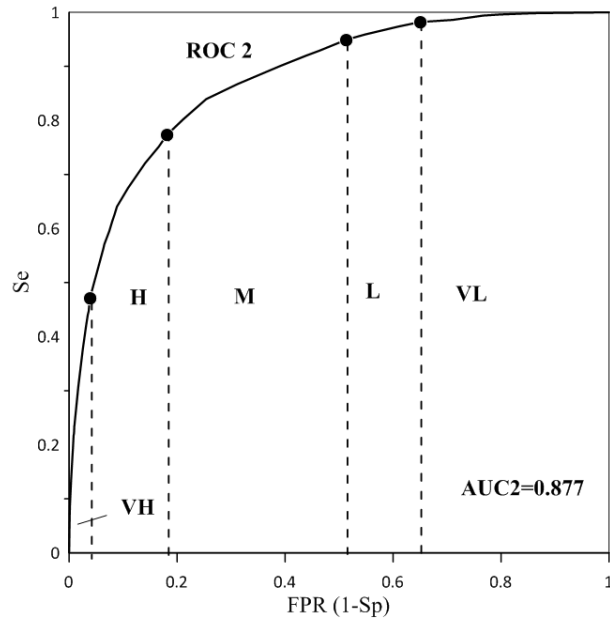


FIGURE B



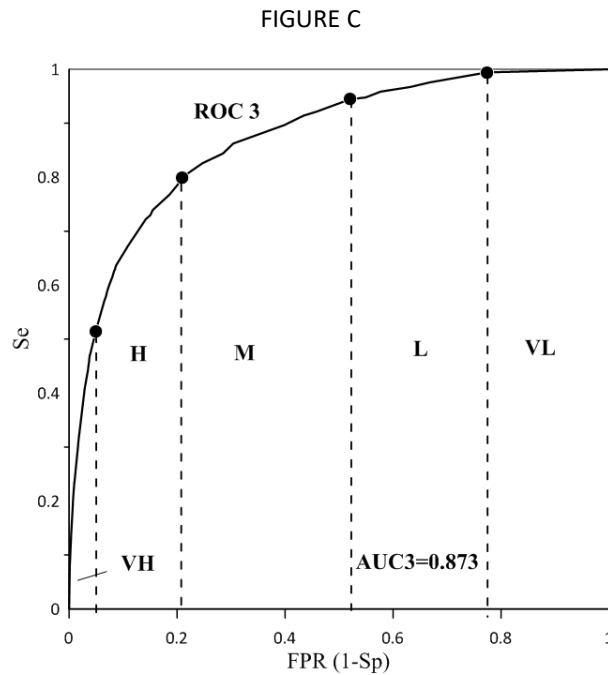


Figure 5 (A, B, C). ROC curves, AUC values and GY class boundaries for the three models

The AUC values for the three models are close to 0.9, and therefore, according to Swets (1988), are highly accurate models that also have a very similar fit. This close similarity between the models' curves, and even in the same method with different parameters, seems to be because the same dataset was used for the three methods, including the valuation methods. The same FR values were used in all the methods to quantify the factors, which could explain more clearly this close similarity. Finally, it should be noted that the present study is designed to compare the classification systems and not the accuracy of the models used.

One property mentioned earlier in the article is the saturation of the AUC value in the horizontal projection. Hence, when low LSI values are truncated – values not included in the map because they do not improve the AUC or have clear stability (e.g., flat zones)– GY presents stable limits with very little variation, especially in the high susceptibility zone. However, for the other classification systems all the levels are greatly altered by this suppression of values because their calculation depends on the overall set of available information.

3.6 LSM classification

Four classification systems were applied: generalised Youden (GY), natural breaks (NB), quantiles (Q) and modified head/tail (HTm). The GY boundaries were calculated using the ROC curve analysis described above (Fig. 5). The NB and Q values were obtained directly from a GIS application, and the HTm are easily calculated from the information provided by the software. Table 8 presents the class boundaries for the four classification types (GY, NB, Q, HTm) in the three maps created for each model.

Table 8. Upper class boundaries for each type and map (LSM1, LSM2, LSM3)

Type	Generalised Youden			Natural Breaks			Quantile			H/T modified		
Class	LSM1	LSM2	LSM3	LSM1	LSM2	LSM3	LSM1	LSM2	LSM3	LSM1	LSM2	LSM3

	GY1	GY2	GY3	NB1	NB2	NB3	Q1	Q2	Q3	HTm1	HTm2	HTm3
VL	35	35	15	50	42	57	31	23	14	40	36	37
L	65	50	40	98	88	147	60	44	33	77*	68*	97*
M	95	86	75	158	151	345	84	66	56	125	123	330
H	149	160	219	228	232	643	113	99	106	190	190	605
VH	340	340	960	337	340	960	337	340	960	337	340	960

(*) Mean of the full series of LSI values for HTm

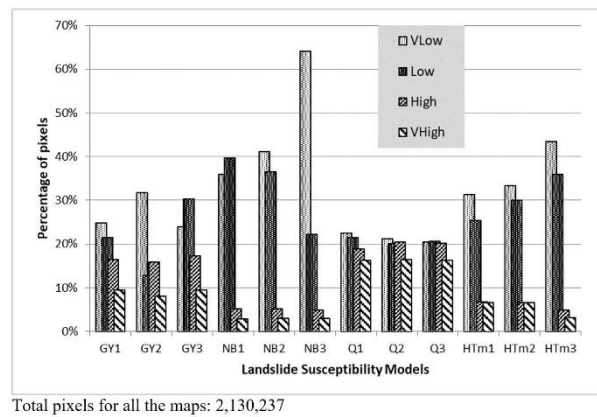


Figure 6. Pixel distribution by class type (except central moderate class) and LSMs (1, 2, 3)

Figure 6 displays the distribution of pixels for the four classification types (GY, NB, Q, HTm) and the three maps obtained for each model. First, a marked difference can be clearly seen in the distribution of pixels among the four classifications. The distribution patterns are more similar when the same model is considered (SMCE 1 and 2). There is also a notable increase in the percentage of pixels (>40%) in NB3 and HTm3 (logit model 3 in the NB and HTm classifications). The Q classes are more homogenous, which is inherent in their definition.

The percentage of susceptible (or potentially unstable) pixels in GY2 is lower than in GY1 (8% in levels VH + H), despite this being an option with the largest weighting in the slope factor. This appears to indicate the importance of the other factors over the slope. Tables 4 and 5 show that weightings above 200 have 40% of the pixels in lithology, 64% in LULC and only 3.2% in slopes. This reduction is inappreciable in the other classifications, however.

4. Comparative analysis of classification and results

The creation of a susceptibility map is concluded when the goodness of fit of the data is confirmed. Importantly, the series of LSMs developed from the same dataset but using different classification methods appears not to show a distribution of discriminant values with the same predictive ability. To test this lack of correspondence, and to obtain a common evaluation of the maps, quantitative indices were used to conduct a comparative analysis of the three LSMs. This comparison identified which maps are similar and which best define landslide susceptibility in the study area. The

four indices used were the stability ratio (Rst), the landslide density index (Rind), the Kappa index (κ) and Moran's index of spatial autocorrelation (MI).

4.1 Stability Ratio (Rst)

In section 3.3 a considerable variability in the number of pixels per class was noted, which could be used to establish a comparison index. Indeed, Fig. 6 shows a large difference between the number of pixels at the high and low levels, which is relevant to learn whether a model is pessimistic or optimistic. Thus, if a large number of cells in the study area reflect high-risk susceptibility levels and, at the same time, relatively few have low-risk levels, the model can be considered pessimistic or conservative. In contrast, if the low levels clearly exceed the number of high levels the model is optimistic or liberal.

One way of expressing this is through the relation of the number of pixels (np) at high levels (VH and H) to that of the low levels (L and VL). In general, the former can be considered susceptible (potentially unstable), whereas the latter are practically non-susceptible (stable). Thus, this stability ratio Rst is defined as:

$$Rst = \frac{npVH + npH}{npL + npVL} \quad [17]$$

Table 6 was used to estimate a mean value of this ratio, assuming the slope factor as the main trigger of landslides. The safe zone thus corresponds to Class 1, with a probability of $P(C1) = 0.002$ of a landslide occurring, and covers 28% of the surface. The unstable zone, Classes 5–8, taken together has a probability of $P(C5-C8) = \Sigma (GFRn)/1000 = 0.91$, and accounts for 13% of the total surface. Therefore, a reasonable Rst would be $13/28 = 0.46$. In conclusion, for the study area a classification with an Rst > 0.46 is pessimistic, and a lower Rst is optimistic.

Table 9. Unstable pixels ratio Rst, with optimistic–pessimistic interpretation

	GY	NB	Q	HTm
LSM1	0.56 (-)	0.11 (++)	0.80 (- -)	0.24 (++)
LSM2	0.54 (-)	0.10 (++)	0.90 (- -)	0.21 (++)
LSM3	0.49 (~)	0.09 (++)	0.89 (- -)	0.10 (++)

(++): very optimistic; (-): pessimistic; (- -): very pessimistic; (~): neutral

Table 9 shows a considerable difference among classifications, although they are quite similar within the same type (with the exception of HTm in LSM3), demonstrating the crucial importance of the classification method. Thus, in all cases the most pessimistic classifications result from zoning by quantiles, that is, the classification yielding the highest number of failures (type I error), implying a need to plan more complex – and probably unnecessary – interventions. In contrast, the most optimistic classifications (NB and HTm) would lead to insufficient prevention with a corresponding increased risk.

4.2 Index of relative landslide density (Rind)

Bearing in mind that the LSM value was used to differentiate susceptible and non-susceptible slopes, slope failures may be expected to appear in cells with higher discriminant scores. A relative landslide density index was used to verify this. The index is defined by the ratio of the density of slope failures of a given susceptibility class to the overall slope failure density. The index takes the following form:

$$R_{ind} = 100 \times \frac{n_i/N_i}{\sum n_i/N_i} \quad [18]$$

where n_i is the number of slope failures observed in a susceptibility class, and N_i is the area covered by the cells of this class (Baeza 2001).

Table 10 shows the Rind values for just the two higher levels, VH and H, taken together. These values are more representative as they reduce the difference in the number of pixels at each level, which are particularly numerous at lower levels; the differences are greater if levels are considered separately.

Table 10. Rind values for classes VH and H

	GY	NB	Q	HTm
LSM1	0.891	0.836	0.870	0.872
LSM2	0.895	0.822	0.886	0.857
LSM3	0.885	0.748	0.885	0.813

In general, the four classifications include a high percentage of landslides at higher levels. GY provides high stable values for the three methods, slightly superior to Q, which contains a higher number of pixels. This appears to confirm that the number of pixels in GY is the most suitable, with almost 90% of the correct classifications in the three methods. NB and HTm are the classifications with the lowest number of pixels at high levels, which may penalise the Rind value. However, NB for LSM3 is the only classification that falls below 0.8 in the table.

4.3 Kappa index (κ)

This statistical index is considered to be more robust than a simple calculation of observed proportion of agreement because 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 can be expressed in the following conceptual terms (Landis and Koch 1977):

$$\kappa = \frac{d - q}{N - q} \quad [19]$$

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. When categories are ordered it is advisable to use weighted Kappa (Baeza 2016), and assign different weights to categories so that different levels of agreement can contribute to the Kappa value. Although different weights can be used, in the present study the Kappa with linear weighting was calculated in all cases (Fleiss et al. 2003). The Kappa value penalises the disagreement between maps linearly, from the smallest to the largest.

This Kappa index was used to verify the amount of change in classes when some of the parameters of the original model or the calculation model are modified. Table 11 compares the classes in the set of LSM calculated, and the means of the maps are compared (for later use in a cluster analysis).

Table 11. Kappa values from a comparison of LSMs for the same classification system and their means

	GY	NB	Q	HTm
LSM1 – LSM2	0.915	0.926	0.938	0.926
LSM1 – LSM3	0.923	0.628	0.911	0.641
LSM2 – LSM3	0.848	0.689	0.946	0.707
means				
LSM1	0.919	0.777	0.925	0.784
LSM2	0.882	0.808	0.942	0.817
LSM3	0.886	0.659	0.929	0.674

These data show that the different classification types retain a very similar scheme despite the variation in factor weightings or the difference in the model. Specifically, the classifications of the multi-criteria LSM models are very similar; the main difference is in LSM1 and LSM3, among which only GY and Q maintain high similarity. This confirms the stability of the SMCE model, since a change in the assignment of the factor weightings does not greatly modify the resulting maps.

Table 12 compares the degree of variation in similarity of the four classification types for the same map, showing the differences between the classification systems applied.

Table 12. Kappa values from a comparison of the four classification systems for the same map.

	GY			NB			Q		
	LSM1	LSM2	LSM3	LSM1	LSM2	LSM3	LSM1	LSM2	LSM3
GY	1	1	1						
NB	0.535	0.542	0.327	1	1	1			
Q	0.829	0.770	0.794	0.433	0.372	0.238	1	1	1
HTm	0.773	0.766	0.477	0.726	0.739	0.723	0.648	0.558	0.351

The high values between GY and Q are noteworthy, while the values between HTm and NB are somewhat lower. The NB classification shows the least similarity with the other classification methods. Surprisingly, there is less agreement among the types of classification than among the different models applied, again highlighting the importance of selecting a good classification system for a susceptibility map.

Finally, a further comparison was made with the European ELSUS 25m map (Gunther et al. 2014). This cannot be regarded as a particularly reliable reference since it was created with generalist criteria and low resolution, but it gives an idea of its correspondence with more local, higher resolution maps. The classification system is based on the percentage of cells with landslides in each level (50-25-15-7%), which does not correspond with the ones used in the present study. Its Rst for mountainous zones of $59/22 = 2.68$ reveals it as remarkably pessimistic, probably because of its low pixel resolution.

Table 13. Kappa values from the comparison of ELSUS with the other maps for the same classification system

	GY	NB	Q	HTm
LSM1	0.255	0.168	0.267	0.215
LSM2	0.251	0.160	0.270	0.211
LSM3	0.251	0.119	0.262	0.152

Table 13 shows GY and Q have the best match with the European map, coinciding with the classifications with higher Rst. Finally, HTm, and particularly NB, present a poorer match as they are low Rst classifications, that is, optimistic.

4.4. Moran's index (MI)

Moran's index is a measure of spatial autocorrelation, namely, the degree of aggregation or clustering. It is characterised by a correlation in a signal among nearby locations in space.

$$MI = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2} \quad [20]$$

where N is the number of cells indexed by i and j; X is the population in cells; \bar{X} the mean of X; and w_{ij} is an element of a matrix of spatial distance weights.

This spatial autocorrelation index shows, first, whether there is an aggregation of values or whether its distribution is completely random, and second, it quantitatively demonstrates the degree of grouping of data for each class. In this study, the radius of influence was set to 1000 metres and the bias was mitigated due to the different number of neighbours for each cell when using intervals.

As expected, the results of the autocorrelation for the susceptibility maps are all significant. In addition, working with a dataset that includes large zones with little slope and low susceptibility values prevents important differences in the autocorrelation values. This was the reason for selecting a mountainous zone representing a high percentage of areas with landslides. This zone, defined in section 3.1 and Fig. 3 (5T), includes 25% of the recorded landslides in the entire La Marina district (34.74 km² of a total 130.18 km²). Even so, low and moderate susceptibility levels were not considered for the same reason.

Table 14. MI values for classes VH and VH-H.

	GY		NB		Q		HTm	
	VH	VH-H	VH	VH-H	VH	VH-H	VH	VH-H
LSM1	0.112	0.125	0.070	0.119	0.137	0.133	0.100	0.121
LSM2	0.096	0.129	0.046	0.097	0.127	0.126	0.111	0.111
LSM3	0.109	0.120	0.081	0.087	0.119	0.123	0.094	0.093

In Table 14 the highest correlation indices appear when the maps are classified with GY or Q. Because the classification by quantiles is more pessimistic and has a higher number of pixels at those levels, it shows a higher autocorrelation in LSM1 and LSM2, although in general it does not clearly exceed GY. However, the lower number of pixels penalises HTm and NB more significantly. GY stands out above the others in map LSM3, with a greater data variation interval. Taken overall, therefore, the classification based on the ROC analysis provides the best grouping for a contained number of pixels (it should result in more homogenous susceptibility zoning).

5. Discussion

In the previous section, a set of indices was analysed that provide information on the behaviour of the four classification types in the three proposed scenarios. The following section explores whether there are significant differences within each classification and amongst them.

Hierarchical cluster analysis (HCA), which groups the classifications by proximity, was applied to determine these relationships succinctly. HCA is a multi-variant statistical technique designed to divide a set of elements into clusters such that the elements in each cluster present high internal cohesion and marked isolation from the other groups and elements. Here, the classification types are grouped by neighbourhoods in the maps and, a priori, there should therefore be four types.

The nearest neighbour method is used (although other methods can be applied that differ very little in the studied example), without standardising and with squared Euclidean distance. This analysis used the stability ratio Rst (Table 9), $Rind$ (Table 10), the Kappa mean for two maps (Table 11), Kappa for ELSUS (Table 13) and Moran's Index (Table 14, levels VH-H).

Initially the four above-mentioned clusters were pursued, but the analysis of the resulting dendrogram points to the advisability of forming five clusters. The dendrogram resulting from the cluster analysis is shown in Fig. 7, and the values of the normalised centroids are displayed in Fig. 8. Normalisation for the Rst ratio took into account the optimal theoretical value $Rst = 0.46$, noted in 4.1.

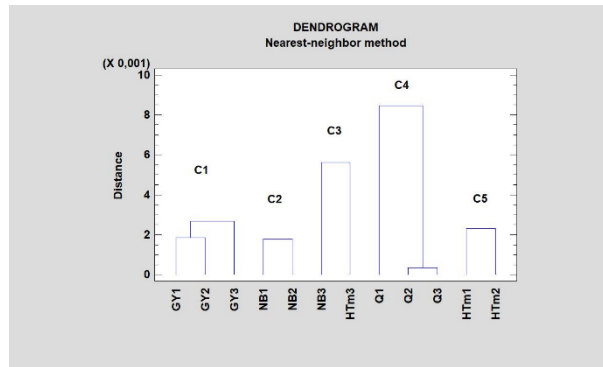


Figure 7. Results of the Hierarchical Cluster Analysis with C1 – C5 clusters

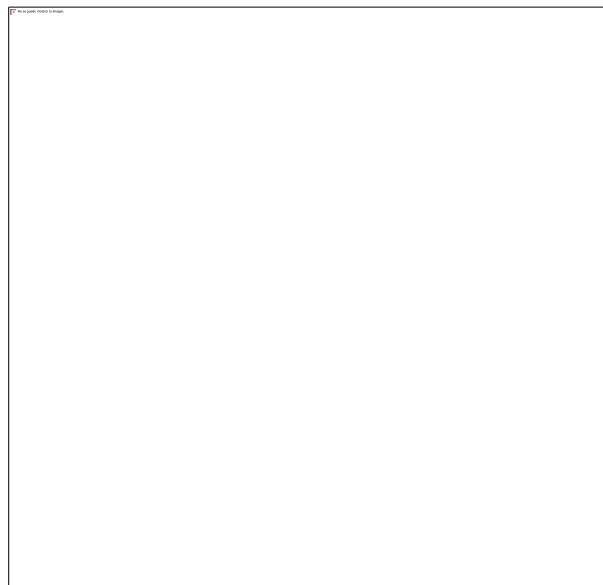


Figure 8. Comparison of cluster centroids

These results are quite illuminating. Notably, the centroids of the cluster that include GY (cluster C1) provide good values for all the indicators considered, C1 being most homogeneous the cluster (see C1 in the dendrogram). The cluster grouping Q (C4) is quite homogenous with good indicator values except for the Rst, which is excessively pessimistic. HTm1 and HTm2 fall within a fairly optimistic cluster (C5), also with good values. Cluster C2, with NB1 and NB2, is very optimistic and has mediocre indices. Finally, the most optimistic cluster (C3) and with the worst values is the one including NB3 and HTm3

GY is the most realistic classification, it registers the change of weights in the slope factor, and considers the internal data structure together with the correct/failed classifications referred to the inventory data, and it is the only system that considers the classifications together with their associated costs. Q is the most pessimistic classification of the four considered and only takes the number of data into account, not the data structure. It is also overdependent on the size of the dataset it is dealing with (if the LSI with low values are removed, all the boundaries fall considerably).

NB takes into account the data structure and is one of the most widely used classification systems in the literature. However, it appears to excessively register the change in exploitation of the original data, and can be considered too

optimistic due to greater variance in the upper range. Our cluster analysis reveals NB as the least suitable classification. Finally, the modified H/T method offers reasonably acceptable results, only penalised by its poor performance in the classification of LSM3.

6. Conclusions

Mapping areas susceptible to landslides is essential for land-use spatial planning and decision making by public bodies responsible for managing affected zones, which are usually located in mountainous areas. Landslide susceptibility is calculated by many types of mathematical models whose degree of accuracy (reliability) needs to be verified; this process can be performed using ROC analysis, among other methods. Classification systems are then applied to the values estimated by these models in order to produce a landslide susceptibility map (LSM) that can be easily applied and interpreted. However, this classification entails a certain loss of information, depending on the criteria adopted in designing the maps, which can seriously alter the original results provided by the models. As Chacon et al. (2006) conclude, "a common standard would be highly valuable for comparing maps, and also to classify landslide areas all around the world".

This study has attempted to exploit all the available information (data on the terrain and from the landslide inventory) to classify an LSM. For the proposed classification, the boundaries derive from the optimal cut-off points calculated by ROC analysis. In addition, unlike most of the literature, the costs of classification errors were also estimated in calculating these boundaries. Only Frattini et al. (2010) have previously used this type of information, but limited to a binary classification.

A rigorous comparison was then made of a set of LSMs for a mountainous region in southeast Spain (La Marina, Alicante) obtained using SMCE and logit models. Various classification systems (GY ROC-based, natural breaks, quantiles and modified head/tail) were applied to the maps to assess their similarities and differences, and their efficiency and consistency in the study area. Several approaches were used (stability ratio, Kappa indices, landslide density index and Moran's spatial autocorrelation index) to compare the classification systems and identify the most suitable ones. Finally, a cluster analysis succinctly grouped and organised the results of these indices.

Consequently, the present study has shown that despite using the same initial data (terrain and inventory) with two different mathematical models, the spatial agreement among the classification systems applied to the susceptibility maps is inconsistent and their spatial patterns differ considerably. The discrepancies among the LSMs are particularly relevant given that overestimation or underestimation of susceptibility at both low and high levels can have major implications for land-use management and civil protection planning. The optimal map should be able to predict most potential landslides and at the same time, efficiently and reliably prevent failures in the study area.

The cluster analysis shows that the maps classified by quantiles (Q) can be considered fairly homogeneous and consistent with GY, but they are excessively pessimistic (a large number of cells in the study are classed as high-risk susceptibility compared with those classified as low risk). These maps provide good results with the comparison indices as they do not take into account the data structure and they are maps with similar distributions. However, in general terms this classification system is not to be recommended since it groups pixels with very different characteristics in the same class, and defines large areas as potentially unstable. Moreover, this classification could motivate plans for more stabilisation interventions, many of which would probably be unnecessary.

In contrast, the classifications based on natural breaks (NB) yield a very heterogeneous result and are also excessively optimistic (a large number of the cells in the study area are classified as low-risk susceptibility). This classification

system is especially affected by large jumps in the data distribution. The HTm classification appears to be fairly balanced although also affected by the data distribution and can lead to over-optimistic results. It has the major advantage that its boundaries are easy to obtain.

The method proposed in the present study (GY) is noteworthy for its good general characteristics. It is the only one that considers the inventory data together with misclassification cost and is therefore the most comprehensive method. On the negative side, it is more complex to calculate, although many programs are available to assist in this process, the most commonly used statistical packages being R (pROC package) and IBM SPSS. The results of this study showed GY to be the most suitable classification method for modelling the terrain susceptibility in the study area.

The poor performance of NB is surprising as it is one of the most widely used and recommended systems (Baeza 2016), but in the present study it is striking for its over-optimistic and unstable classification. This may be due to the data distribution, with large jumps in the tail, which do not correspond with the most appropriate classification levels. It must be stressed that such optimistic maps can lead to loss of life or destruction of infrastructure arising from incorrect classifications of what are actually hazardous zones. By contrast, pessimistic maps that define large unstable zones can lead to a loss of potentially safe spaces or even unnecessary investments in prevention measures in areas that may not represent any hazard. In this latter case, the cost of assuming these errors is lower since human lives are not at risk.

In conclusion, this study has demonstrated that the classification strategy used clearly determines the appearance of results in susceptibility maps, and that classification is more important than the selection of the calculation method. The maps obtained in the study do not have the same meaning and could decisively influence decisions taken in land-use management. There is therefore a need to analyse new classification systems that can be adapted to the needs of the end user so as to guarantee optimal correspondence with the structure of the landslide data in each study area. The method based on ROC analysis provided the best results, and is noteworthy for its consideration of misclassification costs, use of inventory data, homogeneity, stability, and because it is relatively realistic. This research should continue with further testing of the method using other datasets to guarantee its generalised suitability, and it may offer a potential starting point for establishing a standard classification method for susceptibility maps.

BIBLIOGRAPHY

Armstrong, M.P., Xiao, N., Bennett, D.A. (2003). Using genetic algorithms to create multicriteria class intervals for choropleth maps. *Annals of the Association of American Geographers* 93 (3): 595-623. <https://doi.org/10.1111/1467-8306.9303005>

Atkinson, P. and Massari, R., (1998). Generalised Linear Modelling of Susceptibility to Landsliding in the Central Apennines, Italy. *Computers & Geosciences* 24(4):373-385. [https://doi.org/10.1016/S0098-3004\(97\)00117-9](https://doi.org/10.1016/S0098-3004(97)00117-9)

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. <https://doi.org/10.1016/j.geomorph.2004.06.010>

Baeza, C. and Corominas, J. (2001). Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. *Earth Surf. Process. Landforms*, 26: 1251–1263. <https://doi.org/10.1002/esp.263>

- Baeza, C, Lantada, N., Amorim, S. (2016). Statistical and spatial analysis of landslide susceptibility maps with different classification systems. *Environ. Earth Science* 75:1318. <https://doi.org/10.1007/s12665-016-6124-1>
- Basofi, A., Fariza, A., Ahsan A.S., Kamal, I.M. (2015). A Comparison between Natural and Head/Tail Breaks in LSI (Landslide Susceptibility Index) Classification for Landslide Susceptibility Mapping: A Case Study in Ponorogo, East Java, Indonesia. 2015 International Conference on Science in Information Technology, pp 337-342
- Carrara, A., Crosta, G.B., Frattini, P. (2008). Comparing models of debris-flow susceptibility in the alpine environment. *Geomorphology*, 94 (3–4), pp. 353-378. <https://doi.org/10.1016/j.geomorph.2006.10.033>
- COPUT (1998). Lithology, exploitation of industrial rocks and landslide risk in the Valencian Community. Thematic Mapping Series. Department of Public Works of the Valencian Regional Government.
- Chacón, J., Irigaray, C. Fernández, T. and El Hamdouni, R. (2006). Engineering geology maps: landslides and geographical information systems. *Bulletin of Engineering Geology and the Environment*, vol. 65, nº 4, pp. 341-411.
- Chung, C.J.F. and Fabbri, A.G. (2003). Validation of spatial prediction models for landslide hazard mapping. *Nat. Hazards*, 30, pp. 451-472
- Drummond, C. and Holte, R.C. (2006). Cost curves: an improved method for visualizing classifier performance. *Machine Learning*, 65 (1), pp. 95-130
- Duman T. Y., Can, T., Gokceoglu, C., Nefeslioglu, H.A., Sonmez, H. (2006). Application of logistic regression for landslide susceptibility zoning of Cekmece Area, Istanbul, Turkey. *Environmental Geology*, Volume 51, Issue 2, pp 241–256. <https://doi.org/10.1007/s00254-006-0322-1>
- Evans, I.S. (1977). The selection of class intervals. *Transactions of the Institute of British Geographers Vol.2, No.1, Contemporary Cartography*: 98-124. <https://doi.org/10.2307/622195>
- Fleiss, J.L, Levin, B., Paik, M.C. (2003). Statistical methods for rates and proportions, Book Series: Wiley Series in Probability and Statistics. John Wiley & Sons. Print ISBN: 9780471526292. <https://doi.org/10.1002/0471445428>
- Foody, G.M. (2004). Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing* 70 (5): 627-633.
- Fotheringham, A.S., Brunsdon, C., Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*. SAGE Publications, 270 pp

Frattini, P., Crosta, G., Carrara, A. (2010). Techniques for evaluating the performance of landslide susceptibility models. *Engineering Geology*, Volume 111, Issues 1–4, 2010, pp. 62-72, <https://doi.org/10.1016/j.enggeo.2009.12.004>.

Geisser, S. (1998). Comparing two tests used for diagnostic or screening processes. *Statistics Probability Letters*, 40:113–119.

Greiner, M., Pfeiffer, D., Smith, R.D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine* 45, 23-41

Günther, A., Reichenbach, P., Malet, J.P. et al. (2013). Tier-based approaches for landslide susceptibility assessment in Europe. *Landslides* 10: 529. <https://doi.org/10.1007/s10346-012-0349-1>

Günther, A., Van Den Eeckhaut, M., Malet, J.-P., Reichenbach, P., Hervás, J., (2014). Climate-physiographically differentiated Pan-European landslide susceptibility assessment using spatial multi-criteria evaluation and transnational landslide information. *Geomorphology*, 224, 69-85.

Gupta R.P., Kanungo D.P., Arora M.K., Sarkar S. (2008). Approaches for comparative evaluation of raster GIS-based landslide susceptibility zonation maps. *International Journal of Applied Earth Observation and Geoinformation* 10 (3): 330-341. <https://doi.org/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. <https://doi.org/10.1016/j.geomorph.2006.04.007>

Hervás, J. (2017). El inventario de movimientos de ladera de España ALISSA: Metodología y análisis preliminar. In: Alonso, E., Corominas, J., Hürlimann, M. (Eds.), *Taludes 2017. Proc. IX Simposio Nacional sobre Taludes y Laderas Inestables*, Santander, 27-30 June 2017. CIMNE, Barcelona, pp. 629-639.

Jaedicke, C., Van Den Eeckhaut, M., Nadim, F. et al. (2014). Identification of landslide hazard and risk ‘hotspots’ in Europe. *Bull. Eng. Geol. Environ.* 73: 325. <https://doi.org/10.1007/s10064-013-0541-0>

Jenks, G. F. (1967). The data model concept in statistical mapping, *International Yearbook of Cartography* , 7, 186–190

Jiang, B. (2013). Head/Tail Breaks: A New Classification Scheme for Data with a Heavy-Tailed Distribution. *The Professional Geographer*, 482 – 494, vol 65, issue 3. <https://doi.org/10.1080/00330124.2012.700499>

Kiang, M.Y. (2003). A comparative assessment of classification methods. *Decision Support Systems* 35 (4): 441-454. [https://doi.org/10.1016/S0167-9236\(02\)00110-0](https://doi.org/10.1016/S0167-9236(02)00110-0)

Langping L., Hengxing L., Changbao G., Yongshuang Z., Quanwen L., Yuming W. (2017). A modified frequency ratio method for landslide susceptibility assessment” *Landslides*. 14:727 – 741. <https://doi.org/10.1007/s10346-016-0771-x>

Lee. S. (2007). Comparison of landslide susceptibility maps generated through multiple logistic regression for three test areas in Korea. *Earth Surf. Process. Landforms*, 32: 2133–2148. <https://doi.org/10.1002/esp.1517>

Liu, C., Frazier, P., Kumar, L. (2007). Comparative assessment of the measures of thematic classification accuracy. *Remote Sensing of Environment* 107 (4): 606-616. <https://doi.org/10.1016/j.rse.2006.10.010>

López-Ratón, M., Rodríguez-Álvarez, M.X., Cadarso-Suárez, C., Gude-Sampedro, F. (2014). Optimal Cutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *Journal of Statistical Software*, 4, Volume 61, Issue 8.

Malet J.P., Puissant A., Mathieu A., Van Den Eeckhaut M., Fressard M. (2013). Integrating Spatial Multi-criteria Evaluation and Expert Knowledge for Country-Scale Landslide Susceptibility Analysis: Application to France. In: Margottini C., Canuti P., Sassa K. (eds) *Landslide Science and Practice*. Springer, Berlin. https://doi.org/10.1007/978-3-642-31325-7_40

McGee, S. (2002). Simplifying Likelihood Ratios. *J. Gen. Intern. Med.*17: 647-650.

Metz, C. (1978). Basic Principles of ROC Analysis. *Seminars in Nuclear Medicine*, Vol. VIII, No. 4 (October), 183-198.

Nadim, F., Kjekstad, O., Peduzzi P., Herold, C., Jaedicke, C. (2006). Global landslide and avalanche hotspots. *Landslides* 3:159-173. <https://doi.org/10.1007/s10346-006-0036-1>

Ohlmacher, G., Davis, J., (2003). Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology* 69 (3-4): 331-343. [https://doi.org/10.1016/S0013-7952\(03\)00069-3](https://doi.org/10.1016/S0013-7952(03)00069-3)

Powell, R.L., Matzke, N., C. de Souza Jr., Clark, M., Numata I, Hess, L.L., and Roberts D.A. (2004). Sources of error accuracy assessment of thematic land-cover maps in the Brazilian Amazon. *Remote Sensing of Environment* 90 (2): 221-234. <https://doi.org/10.1016/j.rse.2003.12.007>

Saaty, T. (1980). *The Analytic Hierarchy Process*. McGraw Hill, New York.

Smits, P.C., Dellepiane, S.G., Schowengerdt, R.A. (1999). Quality assessment of image classification algorithms for land-cover mapping: a review and proposal for a cost-based approach. *International Journal of Remote Sensing* 20: 1461–1486.

Stehman, S.V., Czaplewski, R.L. (1998). Design and analysis of thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment* (64): 331-344.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science* 240 (4857), 1285–1293.

Van Den Eeckhaut, M, J., Hervás, J., Jaedicke, C., Malet, J.-P., Montanarella, L., Nadim, F. (2012). Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. *Landslides*, 8, pp 357-369.

Varnes, D.J. (1984). *Landslide Hazard Zonation: A Review of Principles and Practice*. Natural Hazards. UNESCO, Paris.

Zhu, X. (2016). *GIS for Environmental Applications*. Routledge, pp 490

Zweig, M.H., and Campbell, G. (1993). Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *CLINICAL CHEMISTRY*, Vol. 39, No. 4, 561-577.