Computers & Geosciences 37 (2011) 410–425

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Evaluation of prediction capability, robustness, and sensitivity in non-linear landslide susceptibility models, Guantánamo, Cuba

C. Melchiorre [a,*], E.A. Castellanos Abella [b], C.J. van Westen [c], M. Matteucci [d]

[a] NGU, Geological Survey of Norway, Leiv Eirikssons vei 39, 7040 Trondheim, Norway
[b] IGP, Institute of Geology and Palaeontology, Blanca 1002 y Carr. Central, 11000 Havana, Cuba
[c] ITC, International Institute for Geoinformation Science and Earth Observation, Hengelosestraat 997514 AE Enschede, The Netherlands
[d] DEI, Politecnico di Milano, Via Ponzio 34/5, 20133 Milan, Italy

ARTICLE INFO

Article history:
Received 27 May 2009
Received in revised form
14 October 2010
Accepted 17 October 2010
Available online 12 November 2010

Keywords:
Landslide susceptibility mapping
Artificial neural networks
Model evaluation

ABSTRACT

This paper describes a procedure for landslide susceptibility assessment based on artificial neural networks, and focuses on the estimation of the prediction capability, robustness, and sensitivity of susceptibility models. The study is carried out in the Guantanamo Province of Cuba, where 186 landslides were mapped using photo-interpretation. Twelve conditioning factors were mapped including geomorphology, geology, soils, landuse, slope angle, slope direction, internal relief, drainage density, distance from roads and faults, rainfall intensity, and ground peak acceleration.

A methodology was used that subdivided the database in 3 subsets. A training set was used for updating the weights. A validation set was used to stop the training procedure when the network started losing generalization capability, and a test set was used to calculate the performance of the network. A 10-fold cross-validation was performed in order to show that the results are repeatable. The prediction capability, the robustness analysis, and the sensitivity analysis were tested on 10 mutually exclusive datasets. The results show that by means of artificial neural networks it is possible to obtain models with high prediction capability and high robustness, and that an exploration of the effect of the individual variables is possible, even if they are considered as a black-box model.

1. Introduction

Since the early eighties many statistical and data-driven approaches have been used to model landslide susceptibility, such as multivariate techniques (Chung et al., 1995; Baeza and Corominas, 2001; Ercanoglu et al., 2004; Komac, 2006), probabilistic methods (Bernknopf et al., 1988; Chung and Fabbri, 1999; Coe et al., 2004; Lee et al., 2004), fuzzy set theory (Davis and Keller, 1997; Chi et al., 2002; Ercanoglu and Gokceoglu, 2004; Kanungo et al., 2006; Saboya et al., 2006), the Dempster–Shafer theory (Binaghi, 1998), and artificial neural networks (Lee et al., 2003a, 2003b; Ercanoglu, 2005; Ermini et al., 2005; Gomez and Kavzoglu, 2005; Lee and Evangelista, 2006; Melchiorre et al., 2008). The availability of a wide range of statistical and data-driven methods makes it difficult to select the optimal method; emphasis should be given to the validation of the model results and on the estimation of the quality of the models. Recently, a comprehensive framework for quality estimation was presented by Guzzetti et al. (2006).

Artificial neural networks (ANNs) are generic non-linear function approximators, extensively used for pattern recognition and classification (Bishop, 1995; Haykin, 1999) which have also a wide applicability in landslide susceptibility modelling. In general, ANNs are presented as a black-box approach and the discussion of the quality of the results is often rather limited, which may be due to the difficulties to apply methods for exploring the behaviour of the networks. Given the potential use of ANN models in risk assessment and planning, it is important to be able to present a clear evaluation of the quality of the models. Furthermore, it is important to better understand how the input variables influence the final susceptibility map in an ANN approach, in order to recognise the main causative factors playing role in landslide occurrence.

This paper presents a framework to estimate the quality of landslide susceptibility models based on ANNs. Methods to evaluate landslide susceptibility models are presented that focus on three main aspects: prediction capability, robustness, and sensitivity. The selected methods are not specific for any statistical or data-driven technique, so they can also be used to compare different models. The methods were implemented for the assessment of landslide susceptibility by means of ANNs in the Guantánamo province in Cuba.

2. Study area, landslide inventory, and causal factors

Guantánamo province (Fig. 1) is located in the east of Cuba, and consists for 75% of mountainous area. The size of the province is

* Corresponding author. Tel.: +47 73904114.
E-mail address: caterina.melchiorre@ngu.no (C. Melchiorre).

0098-3004/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2010.10.004

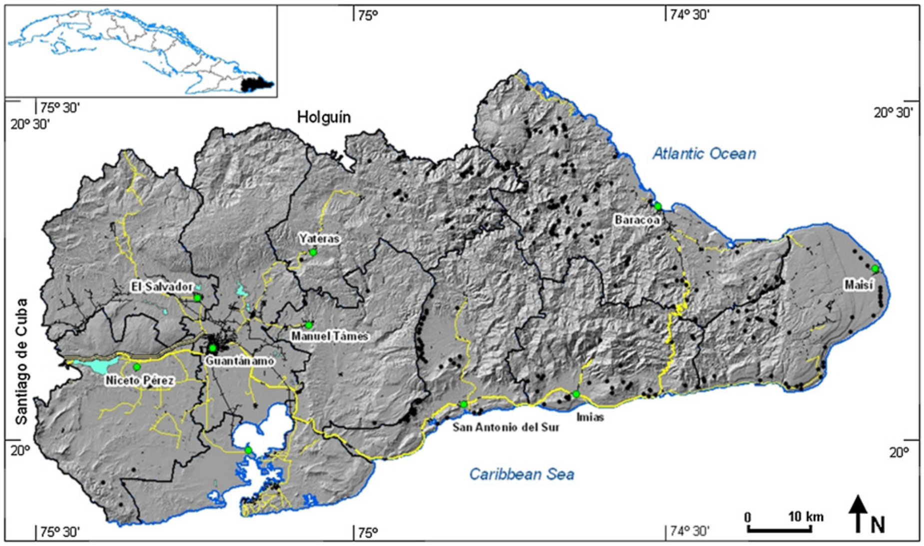

Fig. 1. Location of study area: Guantánamo province, Cuba. The black dots are the landslides.

6186 km^2 and it has approximately half a million of inhabitants (2007). The highest elevation is 1181 m located in the Maisí municipality in the east. Most of the northeast part is mountainous, whereas a large valley, which also forms a separate hydrographic basin, covers the southwest. The northeast basin is drained by the Toa River, which has the highest discharge in Cuba. In terms of climate, Guantánamo contains both the most humid (in the northern) and driest (in the southern) zones of the country. The main geological units are highly weathered ophiolites in the north and metamorphic and sedimentary rocks in the south. The relief is relatively young and strongly affected by tectonic processes.

Guantánamo province is prone to hurricanes: 49 devastating hurricanes were recorded over the period 1789–2003, which occur mostly in the months of September and October. Natural and man-made forest fires are also a major concern. Over the period 1997–2002 93 large fires were reported, affecting an area of 3043 ha. The province is also susceptible to earthquakes due to the proximity of the boundary of the Caribbean and North American plates. Landslides in Guantánamo occur in different geological units but are most frequent in sedimentary rocks.

A landslide inventory map was prepared for the study area, based on stereo airphoto interpretation and field checks. Photo-interpretation was carried out using 300 aerial photos (format 23 mm $\times$ 23 mm) from the year 2000 at 1:25,000 scale covering the entire Guantánamo province. Owing to the availability of only one temporal set of airphotos it was not possible to evaluate the temporal changes in landslide occurence. The temporal information was limited to a few historically known events reported by the local civil defence authorities during fieldwork campaigns. In total, 281 landslides were identified covering an area of about 19.92 km^2. Four main types of landslides were mapped: 22 rockfalls, 26 debris flows, 18 topples, and 215 slides. The slides were subdivided into two sub-groups: 29 large rockslides, located in a high tectonically affected area in the Sierra de Caujerí, and a group of 186 shallow landslides dispersed all over the province. Statistical analysis revealed that the size of landslides varies from 8000 m^2 up to 950,000 m^2 with an average of 70,000 m^2.

For the susceptibility analysis the causal factors were selected based on the literature and on the available data in Cuba. They were divided in 4 groups: ground conditions (i.e., geomorphology, geology, landuse, and soils), morphometric factors (i.e., slope angle, slope direction, internal relief, and drainage density), distance related factors (i.e., distance to roads and distance to faults) and triggering factors (i.e., maximum daily rainfall and peak ground acceleration). A total of 12 maps were produced.

Reflecting several geomorphological processes in the area, the geomorphological map contains 34 types of units including mountains, hills, and plains of different types and geneses. The mountain areas are the most extensive, as they cover 64% of the surface. They were divided into classes according to the genetic process (e.g., tectonic and erosive) and to the relief.

The Guantánamo province has 44 geological units (Fig. 2) from the Mesozoic Ophiolite complex to Holocene surficial deposits. A total of 45.5% of the province is covered by three units; the San Luis Formation with 1051.6 km^2 distributed mainly in the western part, the Sierra del Purial Formation with 907 km^2 mostly located in the east, and the Maquey Formation with 860.1 km^2 mainly outcropping in the central part of province. Each of the other units covers less than 6% of the territory.

Sixteen landuse types cover the study area with predominance of natural forest (55%) and natural pasture (14%). The rest of the territory is covered by sugar cane, coffee, and cultivated pasture and grassland.

The last ground condition factor is soil type, classified by a combination of a group, a sub-group, and parent material. The predominant soil map unit (21% of the area) is the Tropical Greyish Calcareous formed from limestone, marls, and carbonated detrital material. Other predominant soil map units are the "Typical Greyish Tropical Saline", the "Greyish-Red Calcareous", and the "Typical Saline", but they do not cover more than 10% of area. Most of the soils are less than 30 cm deep, which is relevant for landslide occurrence.

The morphometric factors were extracted from a DEM with pixel size of 50 m. The internal relief represents the height difference per unit area. This map (Fig. 3B) was created in ArcGIS by calculating the minimum and maximum elevation per hectare. The drainage density map includes both natural and artificial drainage networks. It represents the length per unit area of drainage lines located in a radius of 1 km. The drainage density map (Fig. 3C) shows concentrations of high values in specific zones, especially in the eastern part of Sierra del Purial and other mountain zones.

Maps of the distance related factors were created for linear features influencing landslide occurrence. We calculated buffer areas around roads (i.e., highways, first and second order roads,

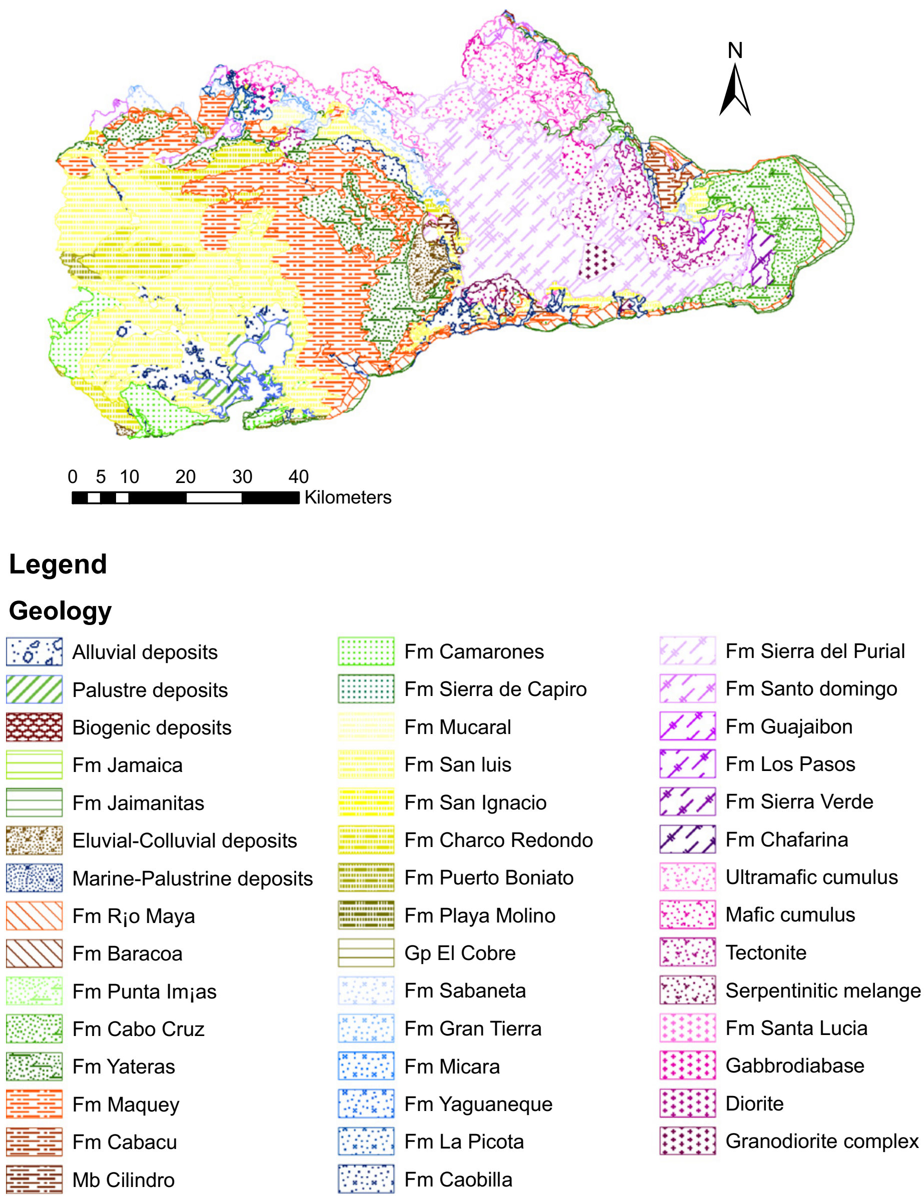

Fig. 2. Geology map.

streets in populated zones, unpaved and enhanced-unpaved roads, trails and tracks, and wide and narrow railways) and faults (inferred faults, certain faults, supposed faults, thrust faults, and reversed faults).

Rainfall and earthquakes were considered as triggering factors and used in the landslide hazard assessment. We used two maps: a raster map of maximum expected rainfall in 24 h for a 100 year return period (Fig. 3E) and a raster map of the peak ground acceleration with 100 years return period (Fig. 3F). Fig. 3E shows a high contrast in rainfall between the northern and southern parts of the province. Rainfall in the northern part, close to Baracoa, usually comes from the northeast (Atlantic region) and is controlled by the relief. The mountains in the central eastern part serve as a barrier for rain clouds and the area south of that is a semi-arid zone. The PGA values (Fig. 3F) are highly influenced by the Caribbean-North American plate boundary located south of the province and by the high seismic activity zone south of Santiago de Cuba city. Intra-plate seismic activity has also been recently detected in the mountainous part of the Guantánamo province.

More information about data acquisition is given in Castellanos Abella (2008).

For each landslide type, the spatial relationship with each factor map was analysed using weights of evidence modelling. Conceptual models were generated for rockfalls, debris flows, topples, and large rockslides (Castellanos Abella, 2008). For shallow landslides it was not possible to generate such a unique conceptual model with bivariate statistical analysis, because they occur under different conditions in the study area.

Given the complex relationship between shallow slides and conditional factors, it was decided to use a non-linear data-driven technique: artificial neural networks.

3. Methodology

Two main tools were used in the modelling: a Geographical Information System (GIS) was used to produce, store, and manipulate the dataset, whereas the actual ANN modelling was implemented in Matlab software.

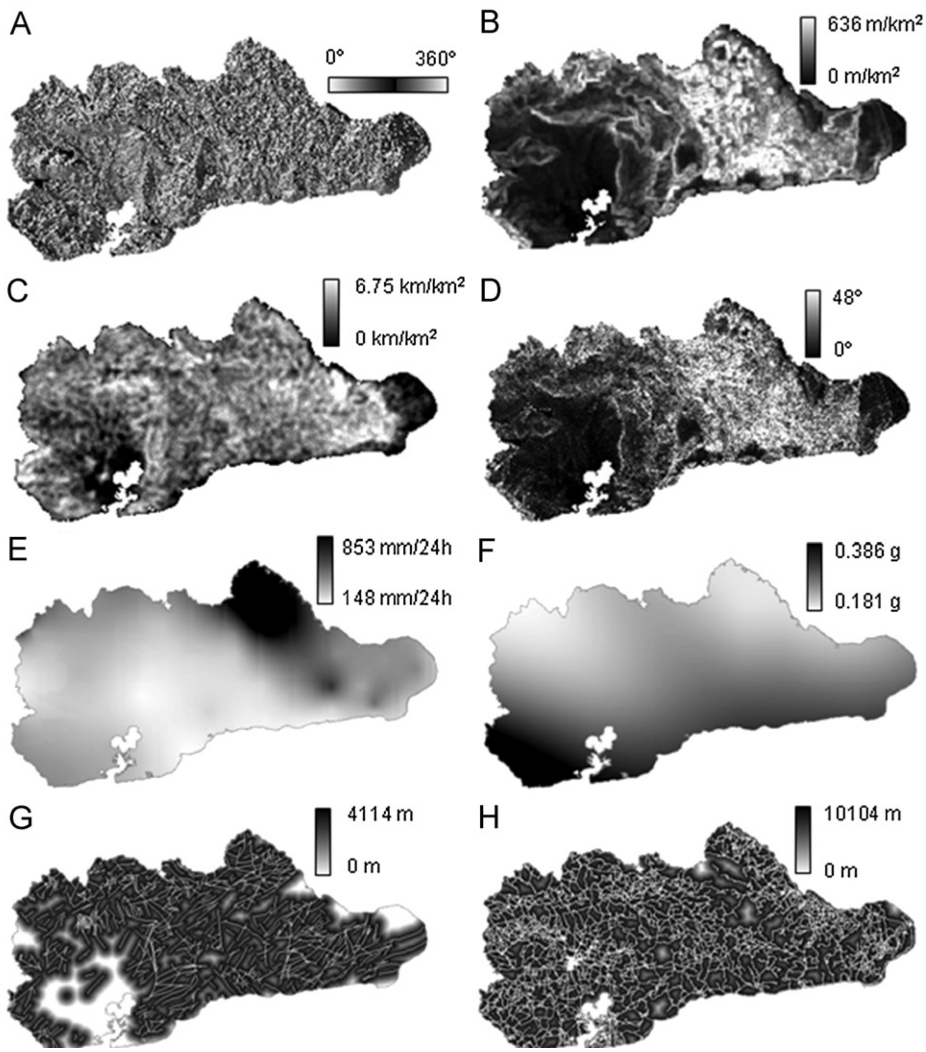

Fig. 3. Some of the causative factors used for susceptibility analysis in Guantánamo province: (A) aspect; (B) internal relief; (C) drainage density; (D) slope angle; (E) rainfall intensity; (F) ground peak acceleration; (G) fault distance and (H) road distance.

3.1. Data selection and data treatment

For the generation of the ANNs the 12 conditioning factors mentioned above were used together with the dataset of 186 shallow landslides.

The selected causal factors (i.e., input variables) were either categorical or continuous datasets. Categorical variables were converted into numerical values by labelling them in a range from 0 to 1. Numerical values were assigned to each class using expert knowledge and considering the influence of the variables on landslide occurrence. The continuous variables were simply scaled in the range from 0 to 1, except for the slope direction (aspect) variable, due to the nature of this data indicating compass direction, ranging from 0° to 360°. To solve this slope direction data was converted into two variables: cosine and sine of the slope direction, which were both scaled in the range between 0 and 1.

The 186 shallow landslides were randomly divided into 10 mutually exclusive subsets, each one containing approximately the same number of landslides. For each subset a number of pixels without landslides equal to the number of pixels with landslides were selected. Since the landslides have different sizes, the total number of pixels in the 10 subsets range from circa 650 to 900, with a pixel size of 50 m.

3.2. Artificial neural networks, training, and cross-validation

A neural network consists of a set of basic units, called neurons. Each neuron computes a non-linear function from its input. Every input has an assigned weight that determines its impact on the overall output of the node. By interconnecting a proper number of nodes in a suitable way and by setting the weights to appropriate values, a neural network can approximate any non-linear function with arbitrary precision (Hornik et al., 1989).

The topology and the weights are the key point in the analysis, since they determine the final behaviour of neural networks and the ability of a network to learn a specific dataset. Fig. 4 shows the steps followed during the analysis. The number of hidden layers was initially set at 1 and a log-sigmoid activation function was chosen for all the network connections. Those choices allow modelling any non-linear continuous functions (Bishop, 1995). The Levenberg–Marquardt algorithm (Hagan and Menhaj, 1994; Marquardt, 1963) was selected among the training algorithms, since it allows high-speed convergence without requiring any tuning in the learning parameters. The training procedure was performed using the early stopping criterion (Caruana et al., 2000), in order to avoid overfitting (i.e., modelling errors and noise from the data) and to guarantee generalization capability. The early

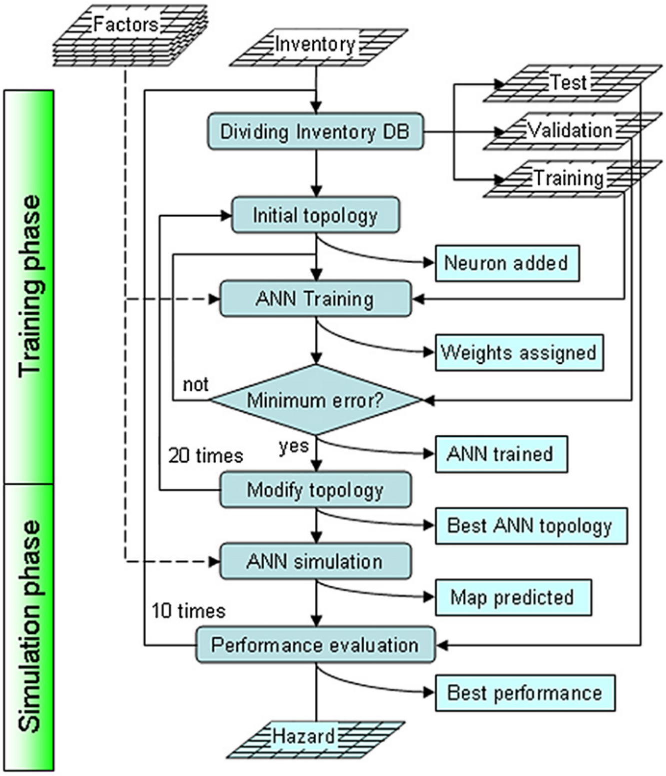

Fig. 4. Scheme of the procedure used to train the networks.

stopping criterion is based on the use of three datasets (i.e., training, validation, and test set). A training set was used for updating the weights. A validation set was used to stop the training procedure when the network started loosing generalization capability (i.e., when the error on the validation set starts to increase), and a test set was used to calculate the performance of the networks. In order to select the best network topology, the number of hidden neurons was progressively increased from 1 to 20 neurons, always using the early stopping criterion to end the training. That means that for each tested topology the training was stopped when the error on the validation set reached its minimum. The final number of hidden neurons was chosen according to the performance of the network on the test set.

A 10-fold cross-validation was performed in order to show that the results are repeatable and not dependent on a specific subsample of the database. Iteratively, 8 subsets were used to train the network, 1 as validation set to stop the training, and 1 as test set to evaluate the performance of the network. Since the procedure was repeated 10 times, that means that each of the 10 subsets was used once as test set. For each of the 10 subsamples of the database, the optimal topology was found by incrementing the number of neurons from 1 to 20 and ending the training with the early-stop criterion, as previously described. In total we obtained 10 best trained networks and so 10 susceptibility models on which we evaluated the prediction capability and performed robustness and sensitivity analysis.

3.3. Prediction capability

The prediction capability of the model was evaluated to estimate the ability to forecast unknown events (i.e., events not used during the fitting and the building phase of the model). When the landslide database can be split in temporal landslide inventory maps, the analysis of the temporal prediction capability is also possible. Since for the Guantanamo area a temporal subdivision of the landslide inventory was not possible, due to the lack of multi-temporal images, only the spatial prediction capability was evaluated.

In the susceptibility modelling a subset of the entire study area, including all landslides (186 for a total of 3087 pixels) and only part of the area without landslides (3100 pixels) was used to select the best models (in this case to find the networks with the highest performance on the test sets). Since we used 10-fold cross-validation, 10 models were fitted and then used to produce the susceptibility maps for the entire study area. Therefore, when modelling landslide susceptibility, it is necessary to evaluate the accuracy of the classification also for the obtained susceptibility maps. At first, the prediction capability of the trained networks was evaluated on the test sets and then for the entire study area. Basically, in the first step a threshold in the output of the network was used (i.e., 0.5) and the capability to separate the class of unstable areas (i.e., output $\geq$ 0.5) from the class of stable areas (i.e., output $<$ 0.5) was evaluated. In the second step the output of the network was used in a continuous way and ranked in a number of fixed classes with equal size. Then the prediction capability of each class was measured.

The prediction rate of the trained networks was estimated by three performance measures: sensitivity measure, specificity, and overall accuracy (Altman and Bland, 1994; Duda et al., 2000). The sensitivity measure is the percentage of correctly classified landslides, the specificity is the percentage of correctly classified landslide free area, and the overall accuracy is the percentage of correctly classified cases (both landslides and landslide free areas). Since an equal number of landslide free areas and landslide areas were selected, the overall accuracy is the mean value of sensitivity measure and specificity. The results of those performance measures were calculated for each subdivision used in the cross-validation.

The prediction capability of the susceptibility maps was estimated using the prediction rate curve (Chung and Fabbri, 1999, 2003). This curve shows the cumulative percentage of correctly classified landslide areas versus the area of the map with highest susceptibility classes. When calculating this percentage for the test set, the curve shows the capability of the model to predict landslides and so the prediction capability of each susceptibility class. It is possible to assign a measure of prediction power to each susceptibility class by comparing the probability that a pixel is affected by landslides in the class and the probability in the whole study area (Chung and Fabbri, 2003). This is equivalent to calculating the ratio between the landslide area in each susceptibility class and the number of susceptibility classes (Guzzetti et al., 2006) when all the susceptibility classes have the same size. Although the value of the index can vary widely for different study areas and high index values are difficult to obtain in complex areas (Guzzetti et al., 2006), the index can be used as an indicator of the goodness of prediction in each susceptibility class for a given study area and for comparison of models.

3.4. Robustness analysis

The term robustness analysis is in general used to evaluate the change in the accuracy of the classification due to perturbation in the computational flow (Alippi et al., 2004). In this paper we only focused on disturbance due to errors in input data. The idea of this analysis was to evaluate how errors in the input data can affect the assessment of landslide susceptibility with respect to possible changes in model performance. The change in performance was evaluated after changing the input data according to an error model. In this study the robustness index proposed by Alippi et al. (2004) was applied in its modified version as used in Melchiorre et al. (2006).

Since the correct classification of the area with landslides is the final aim of the susceptibility analysis, the sensitivity measure was selected to evaluate the robustness index. The basic idea was to quantify how the rate of classification of unstable area (i.e., sensitivity measure) changed due to changes of the conditioning factors. The robustness index U is defined as the difference in sensitivity measure due to a generic Δ variation in the input data:

$$U(\Delta,\theta) = |Sen(f(x_\Delta,\theta)) - Sen(f(x,\theta))|$$

where $f(x,\theta)$ and $f(x_\Delta,\theta)$ are the answers of the network with not varied and varied inputs, respectively.

In order to evaluate the impact of the errors, a robustness index $\overline{\gamma}$ is introduced. The network is robust at level $\overline{\gamma}$ in D, when $\overline{\gamma}$ is the minimum positive value for which

$$U(\Delta,\theta) \le \overline{\gamma}, \ \forall \Delta \in D, \ \forall \gamma \ge \overline{\gamma}$$

For more details on the calculation of $\overline{\gamma}$, refer to Alippi et al. (2004).

Regarding the error model, we adopted a conservative approach by sampling the perturbed values from a uniform distribution with a maximum error equal to 5% of the measure.

3.5. Sensitivity analysis

The main task in landslide susceptibility assessment is to find out how the causal factors influence the occurrence of landslides. This can be a first preliminary check of the model from a geomorphological point of view. It might be that a data-driven model predicts the occurrence of landslides, but with causal factors without any clear explanation with respect to the physical process of landsliding: the model is reproducing the presence of landslides, but does not give a plausible geological mechanism for the landslide occurrence.

With the term sensitivity analysis we indicate the investigation of the behaviour of the networks, in terms of how each input variable influences the final output. Since neural networks are non-linear systems, the detection of their behaviour in their domain could be quite complex. Therefore a graphical method (Plate et al., 2000), already successfully used in practical applications by Cannon and McKendry (2002), was used to analyse the function computed by a Multi-Layer Perceptron (MLP) network. This approach allows qualitatively investigating the effect of conditioning factors on the models.

The idea of this graphic method is to represent in a scatter plot the contribution of each input variable to the final output at certain points in its domain (e.g., training sets and random points). The input variables are plotted on the x-axis and the variations (Δ_i) of the output on the y-axis. The Δ_i values represented the variations of the model from an arbitrary baseline (b_i) to their original values and they were calculated as

$$\Delta_i = Y(X) - Y(X_1,\ldots,X_{i-1},b_i,X_{i+1},\ldots,X_k)$$

where X are the input variables and Y are the model outputs. By definition, when X_i is equal to the baseline b_i, Δ_i is zero. Instead of plotting the variation of model output using points, as in standard scatter plots, effects are plotted as small segments, with slope equal to the partial derivative of the model output with respect to X_i. The visualization of the partial derivatives as segments allows identification of trends and the type of non-linear relationships between each input variable and the model output.

Combining in the same plot information on X_i, Δ_i, and local derivatives allows for studying trends, interactions between variables, non-linear and additive effects of the function computed by the neural networks. The overall Δ_i range measures the importance of the variable, whereas the vertical Δ_i range at points indicates interactions of the variable with other inputs. Trends and non-linearity visible in the sensitivity plots are directly related to trends and non-linearity of the model.

By visually interpreting the sensitivity plots, we can evaluate the following:

(1) The effect of input variables on the output. Variables with no effect on the model appear as horizontal lines.
(2) The variable importance, described by the overall Δ_i vertical range. The greater the overall Δ_i vertical range, the greater the influence of the variable on the model.
(3) The interaction with other variables, described by the spread of Δ_i along the y-axis. Variables with no interaction appear as single lines.
(4) Trends and non-linearity shown by trends and non-linearity of the derivatives. If the small segments describe curves, the function computed by the network is not linear.

3.6. Influences on the spatial prediction pattern—empirical method

As the last step of the analysis, an empirical method was applied to qualitatively investigate the robustness of the final susceptibility map and its sensitivity to each causal factor.

For the robustness estimation, all the input variables were changed by sampling the values from a uniform distribution with maximum error equal to 5% of the measure. The change of prediction capability was then evaluated for the susceptibility map obtained with this perturbed model and compared with the original model (i.e., the not perturbed one).

The estimation of the effect of each casual factor on the spatial susceptibility pattern was done by substituting each variable at the time with completely randomized numbers. The effect was evaluated by comparing the susceptibility maps so produced and their prediction rate curves with the original ones. This approach is based on the assumption that the more a variable is influencing the landslide susceptibility, the more the spatial landslide pattern will be modified in its absence.

4. Results and discussion

All the results presented here were acquired by applying the methods described in Section 3 on the 10 mutually exclusive test

Table 1
Results of the performance measurements for the 10-fold cross-validation.

Subdivisions	Number of neurons	Training sets			Test sets		
		Performance			Performance		
		Sen	Spe	Acc	Sen	Spe	Acc
1	11	88.9	81.1	85.2	86.2	82.9	84.3
2	12	95.9	74.2	84.8	89.5	70.5	78.9
3	16	87.7	75.4	81.6	91.7	76.4	84.7
4	11	87.9	73.1	80.6	86.3	72.5	79.1
5	8	84.5	72.0	78.1	72.2	73.9	72.9
6	8	81.2	78.3	79.8	86.1	73.7	79.3
7	13	89.6	78.9	84.2	87.3	82.9	84.9
8	10	89.6	80.4	85.0	82.2	83.9	83.1
9	14	95.9	84.9	90.6	89.6	82.8	85.5
10	10	90.4	81.5	86.0	85.0	82.4	83.7
	Min	81.2	72.0	78.1	72.2	70.5	72.9
	Max	95.9	84.9	90.6	91.7	83.9	85.5
	Mean	89.2	78.0	83.6	85.6	78.2	81.6
	Std. dev.	4.5	4.2	3.6	5.4	5.3	4.0

sets of the 10-fold cross-validation. Regarding the empirical method only the results for 1 of the 10 models will be presented. We refer to subdivision 1, 2,…,10 for the 10 mutually exclusive test sets of the 10-fold cross-validation, respectively.

4.1. Prediction capability

The performance measures shown in Table 1 show quite good prediction capability for the test set, especially for the landslide subset. If we exclude the results for the subdivision number 5, the sensitivity measure is always higher than 80%, guaranteeing a good classification of the unstable areas. The mean value for the sensitivity measure is 85.6. Since the 10 sets are mutually exclusive, that means 85.6% of the total number of landslides was correctly classified. The standard deviation equal to 5.4 shows that there is a spread in the estimation of the sensitivity measure, but this spread is reasonable and acceptable. Slightly poorer results were obtained for the classification of the stable area.

The following step in the evaluation of the accuracy of the results was to estimate the prediction of the susceptibility maps. Each of the trained 10 networks was simulated on the whole study area and the prediction rate curves were calculated. Again we are referring to the prediction rate, since the curve was calculating only on the test set. The curves in Fig. 5 evidence high prediction capability for the most susceptible classes. For instance for the curve labelled as subdivision 9, 10% of the most susceptible area predicts almost the 80% of the landslides in the test set. The worst result was obtained again for subdivision 5: a low overall accuracy (as shown in Table 1) has as consequence a poor discrimination between stable and unstable areas and an overestimation of unstable area in the final susceptibility map (as shown in Fig.5).

Analysing in more details the curve obtained with subdivision 9 and comparing the prediction rate curve with the success rate curve (Fig. 6), it is evident that the two curves do not show significant differences. Since we used a 10-fold cross-validation, explained in more detail in Section 3.2, the prediction rate curve is calculated on circa 10% of the landslides, whereas the success rate

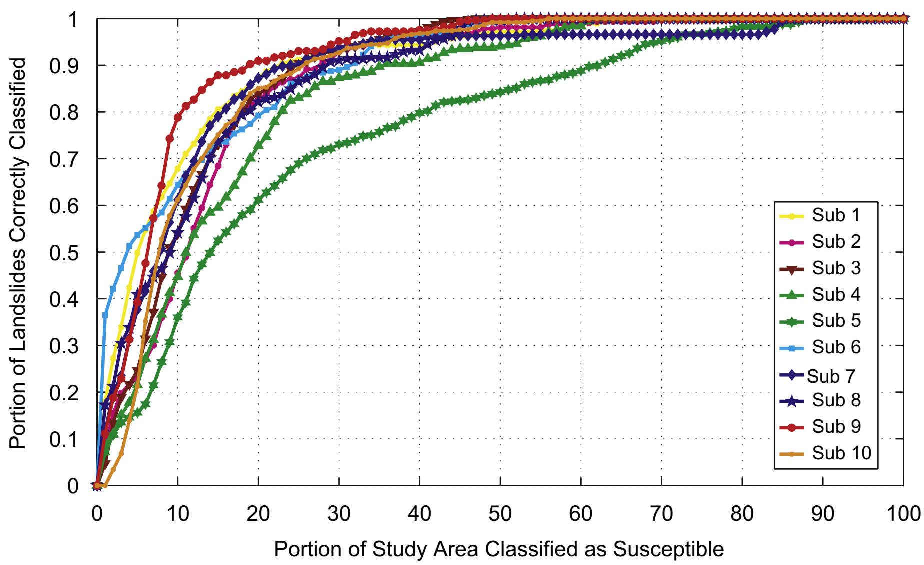

Fig. 5. Prediction rate curve for the 10 models.

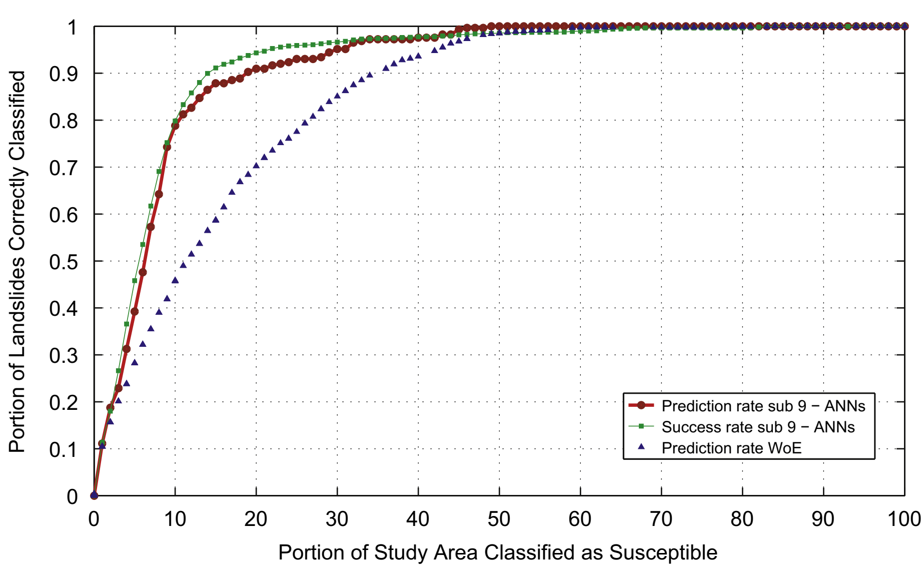

Fig. 6. Prediction and success rate curve for subdivision 9, compared with prediction rate obtained by means of the Weight of Evidence method.

curve on the whole set of landslides. For the most susceptible class (10%) the percentage of correctly classified landslides is 80% for both the success and the prediction rate curves. This proves the high generalization capability of the trained networks.

In order to emphasize that it is a very good result, this curve was compared with the ones calculated for susceptibility models obtained with a Bayesian method (Castellanos Abella and van Westen, 2008). Fig. 6 shows the higher prediction capability of the susceptibility models obtained by means of ANNs. The two worst results achieved by means of ANNs (i.e., subdivisions 4 and 5) are comparable with the best results achieved by using the Bayesian method. We want to underline that the comparison between the curves in Fig. 6 is not rigorous, since they should have been calculated on the same subdivision of the dataset. This was not applied, since the analysis by means of the Weight of Evidence was carried out in a separate work (Castellanos Abella and van Westen, 2008).

Since each of the 10 best trained networks represents just one sample of the obtained failure models, we produced two maps as final products, one representing the mean value of the outputs from the 10 networks (Fig. 7a) and the second one their standard deviation (Fig. 7b). The idea behind this approach is that it should be possible to delineate areas with high susceptibility and low susceptibility only if the 10 models give similar responses, so only where the standard deviation is low. Since this averaged model was produced by using all the landslides of the dataset, only the calculation of the success rate curve is possible. The curve is shown in Fig. 8 and represents the accuracy of the final susceptibility map. Although the prediction rates, estimated with the 10-fold cross-validation, are always higher or comparable to the ones obtained by Weight of Evidence, the variability of the results is high, as shown in Fig. 7b. This is especially true for susceptibility values around 0.5, whereas values close to the extremes 0 and 1 (stable and unstable areas, respectively) have low standard deviation. The simultaneous

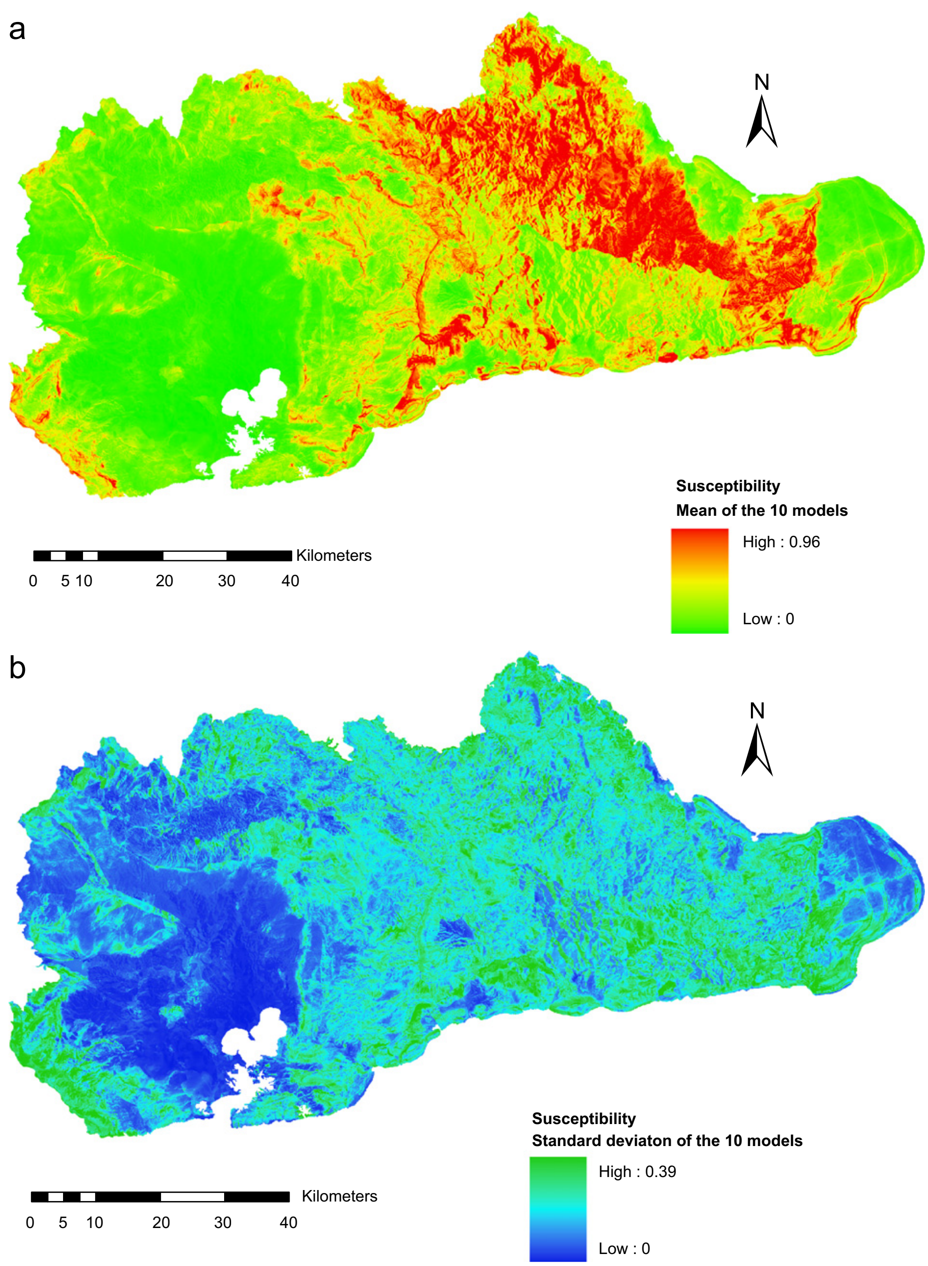

Fig. 7. Mean (a) and standard deviation (b) of the susceptibility models.

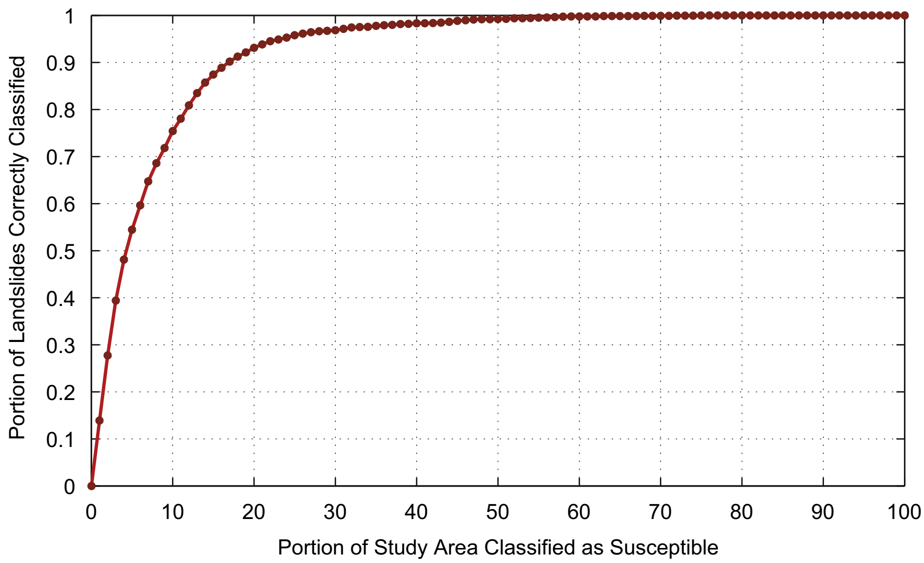

Fig. 8. Success rate curve for the average model.

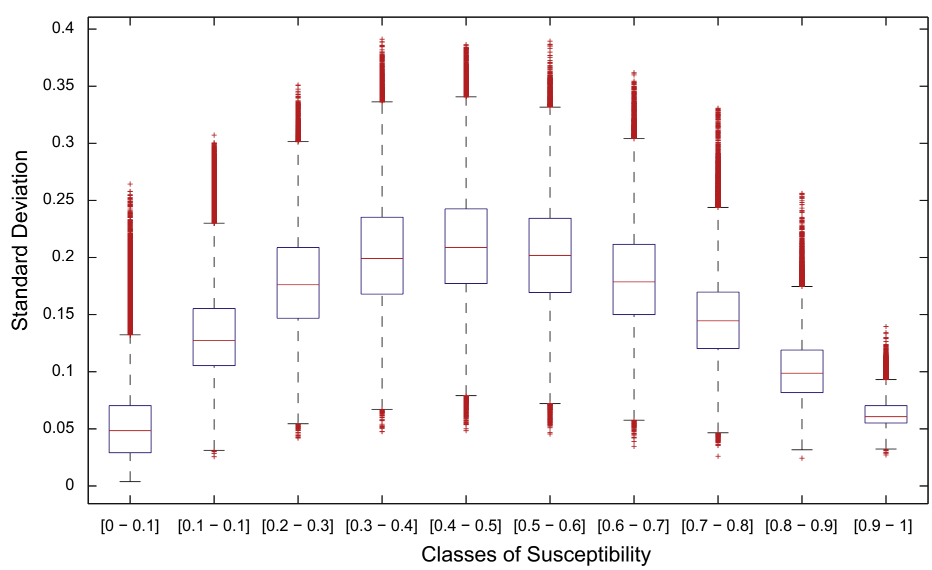

Fig. 9. Boxplot of the standard deviation in each susceptibility class.

Table 2
Results of the robustness indices.

	Subdivisions										**Mean**	**Std. dev.**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**		
Cos aspect	0.4	0.3	0.8	1.1	0.2	0.3	0.6	0.9	0.7	0.0	0.5	0.4
Sin aspect	0.7	1.2	0.2	0.0	0.4	1.5	0.8	1.5	0.7	0.3	0.7	0.5
Drainage	2.1	1.2	0.8	0.8	0.5	0.9	1.7	1.8	1.0	1.8	**1.3**	0.5
Distance to faults	0.4	0.6	0.4	0.3	0.2	0.0	0.8	0.9	0.3	0.0	0.4	0.3
Geology	1.1	0.3	0.8	0.5	0.0	1.2	0.8	1.2	0.7	0.5	0.7	0.4
Geomorphology	0.7	0.9	0.4	1.1	0.4	1.2	0.6	1.8	0.7	0.3	0.8	0.5
Internal relief	1.8	0.9	0.4	1.1	1.1	1.5	0.3	1.8	0.7	1.3	**1.1**	0.5
Landuse	0.4	0.3	0.2	0.8	0.2	2.1	0.0	0.6	0.7	0.3	0.6	0.6
Pga	1.8	1.5	0.4	1.1	0.9	2.1	0.8	1.5	1.0	0.3	**1.1**	0.6
Rainfall	1.1	1.5	1.1	1.1	0.4	1.5	0.3	1.2	0.7	0.8	**1.0**	0.4
Distance to roads	0.7	0.6	0.6	0.3	0.2	0.9	0.0	0.6	0.3	0.0	0.4	0.3
Slope angle	1.1	0.9	0.6	1.1	0.5	1.2	0.8	1.5	0.7	0.8	**0.9**	0.3
Soil	1.4	0.9	0.4	1.3	0.2	1.8	0.3	0.9	1.0	0.3	0.9	0.6
Index tot	3.2	2.2	1.7	2.2	1.6	3.3	2.3	3.1	2.1	2.6	**2.4**	0.6

use of the mean value and of the standard deviation allows for estimating the confidence for each defined susceptibility class. In the case of Guantánamo, both the high and low susceptible classes are estimated with high precision, whereas a higher uncertainty is present for the other classes (Fig. 9).

4.2. Robustness analysis

Also this analysis was done on each of the 10 sets of the cross-validation. For each variable the error was inserted as explained in Section 3.5 and the change in the prediction capability was

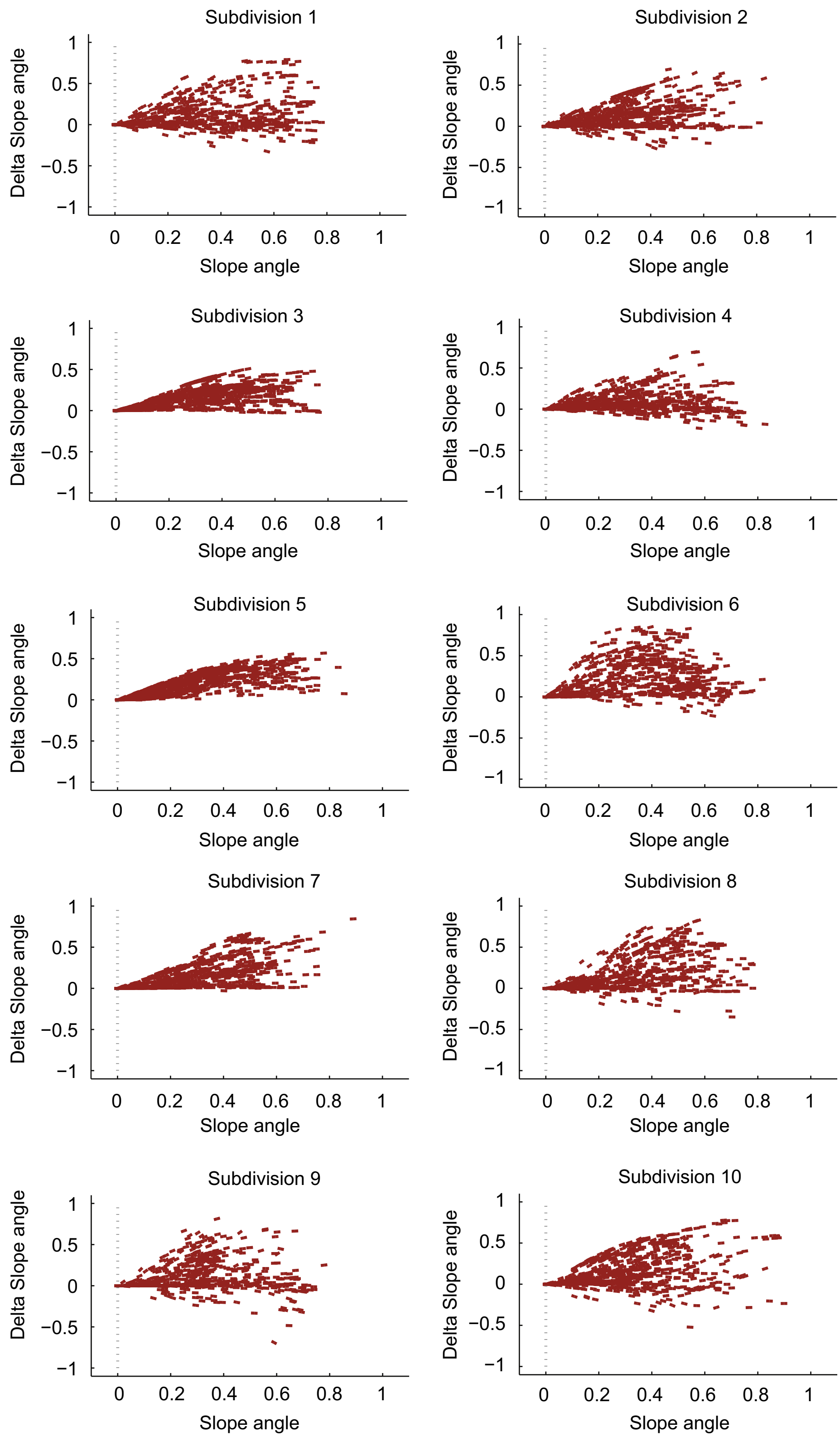

Fig. 10. Sensitivity of the model to the variable "slope" in the 10 subdivisions of the database.

evaluated by measuring the change in sensitivity. The results of the robustness analysis in Table 2 evidence a quite low value of the robustness index. The maximum value reached is 2.1 for the variable drainage in subdivision 1, meaning that the sensitivity is estimated to change 2.1 points when the error is inserted. Considering that the sensitivity measure is 86.2 (see Table 1), the estimated change of sensitivity measure due to perturbation of the input variables is considered acceptable. This indicates that the perturbation in the input does not affect the trained network significantly, and that the obtained models are robust. The index

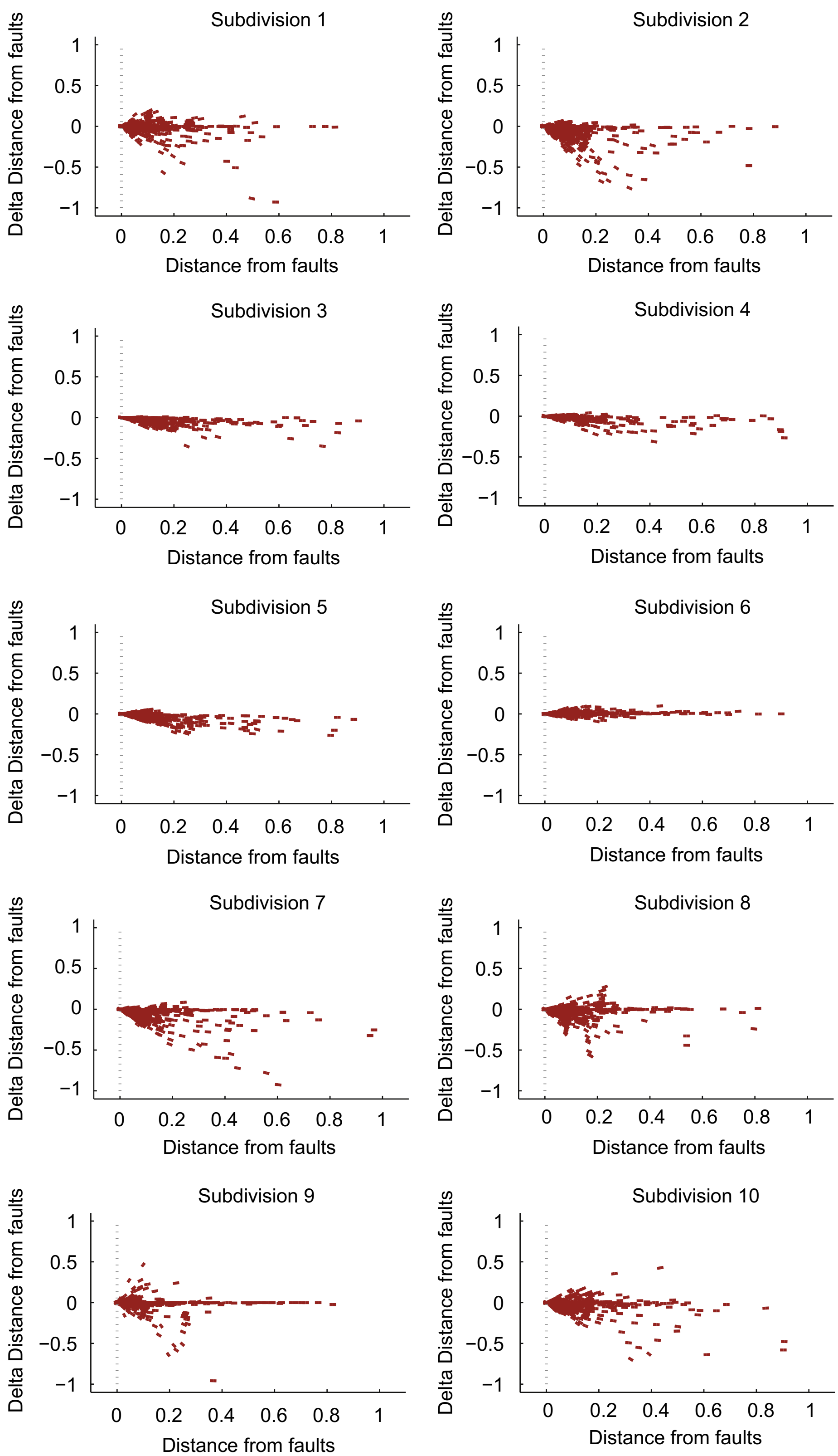

Fig. 11. Sensitivity of the model to the variable "distance to faults" in the 10 subdivisions of the database.

values do not show significance difference among the variables. The highest values of the mean were obtained for drainage, internal relief, and peak ground acceleration (PGA). According to the definition of robustness index previously given, those are the variables in which the inserted errors had a larger influence on the computational flow. The robustness index could be used to delete variables which do not contribute to the performance of the model, but which contain only noise. This was not done for the analysis in Guantanamo, since the values of the robustness indices were low for all variables. If we examine (Table 2) the value of the indexes obtained when all the variables are perturbed, it is possible to notice that the overall robustness of the model is confirmed: the change of sensitivity measure is estimated at 2.4 points in average. This means that even assuming a relative high error (e.g., 5%) simultaneously inserted in all the causal factors, only a mean change of 2.4 point in sensitivity measure is estimated.

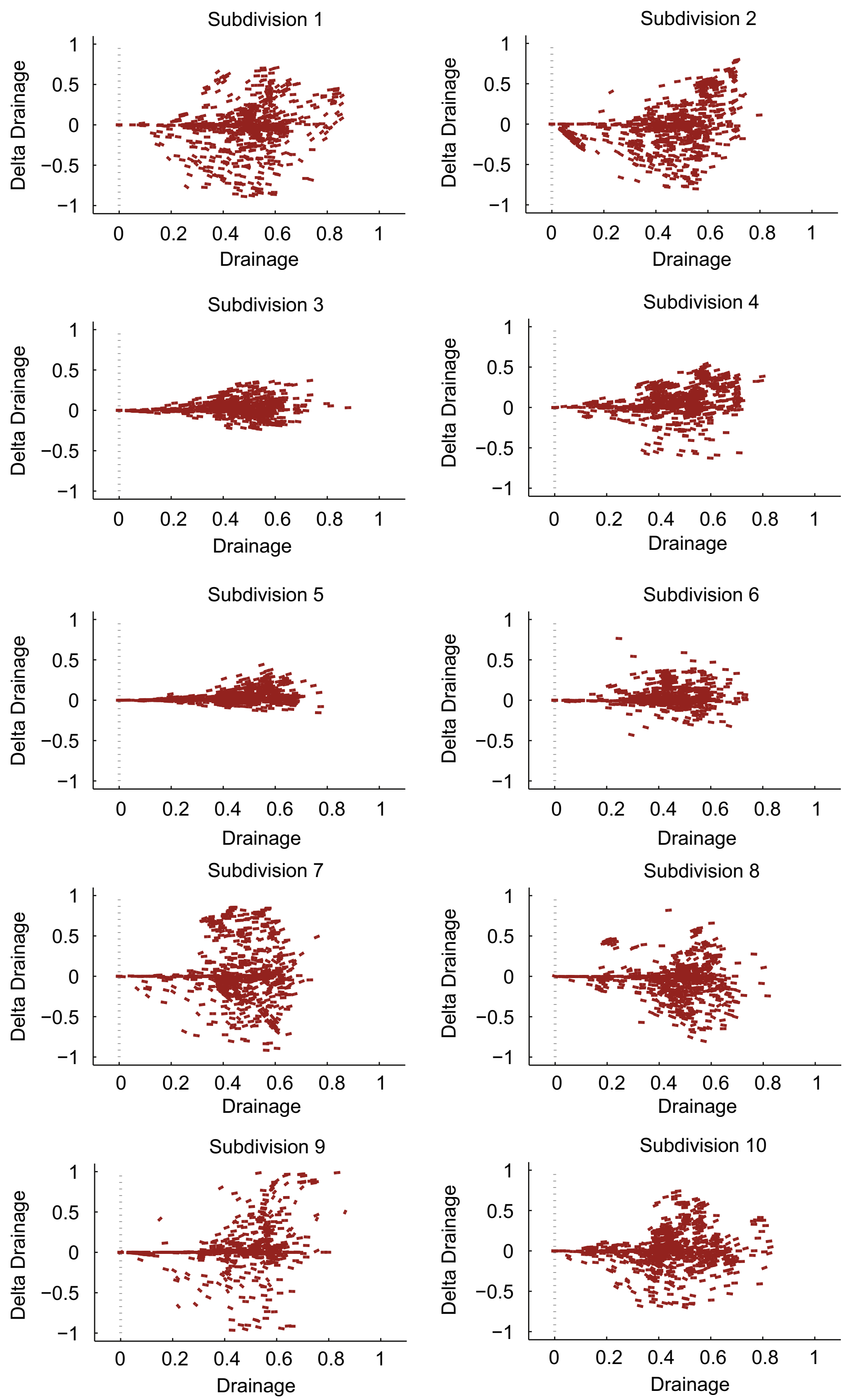

Fig. 12. Sensitivity of the model to the variable "drainage" in the 10 subdivisions of the database.

4.3. Sensitivity analysis

This part discusses the results of the sensitivity analysis applied on the 10 models. As explained earlier, the aim of this analysis was to detect trends, interactions between variables, and non-linear behaviour. This graphical method allows a deeper understanding of the models, especially when the users want to recognise the effect of each causal factor and of the combination of them on the susceptibility to landslides.

In general, no pure linear or non-linear behaviour was detected in the analysis, but all the variables interacted with others. Each variable was detected to have approximately the same behaviour on the 10 models used in the cross-validation, but some variations from the average behaviour were identified. Going into the details of the results, a positive influence on the network for the following variables was detected: slope, internal relief, rainfall, and peak ground acceleration. This means that a positive increment of those variables is associated with a positive increment of the network output and so of the susceptibility level. This means that the model will give an increase of the susceptibility level when slope angle, rainfall, peak ground acceleration, and internal relief increase, which is logically explained from a process point of view.

In Fig. 10 the sensitivity graphs for the slope angle factor are shown. A linear positive influence is mainly detected in the 10 models, visible for example in subdivision 5, even if a non-linear behaviour was found for subdivision 6. The vertical spread in the plots proves the presence of interaction. Since the degree of vertical spread is a qualitative measurement of the variable importance, we can conclude that slope angle has a high influence on the susceptibility model.

An opposite trend was indentified in the same subdivision for the variables "distance to roads" and "distance to faults". An increase of distance results in a decrease in landslide susceptibility. Again, the same behaviour was not recognised in all the subdivisions. As shown in Fig. 11 for the variable "distance to faults" most of the models shows a negative influence on landslide occurrence, but some models show no unique trend. If we compare the degree of vertical variability of the variable "distance to faults" with the variable "slope angle", we can observe that the strength of influence of former is lower than the latter.

For other variables no unique trends could be identified, but since they show interactions, we can conclude that their positive or negative effect on the model depends on the values of other variable (i.e., it depends on the position in the multidimensional space).

For only one variable a non-linear behaviour was identified in the majority of the 10 models. This variable is the "drainage density" for which the graphs of the sensitivity are shown in Fig. 12. Observing the plots, it is possible to notice that in most of the models the drainage pattern has a high influence on landslide susceptibility, but in subdivisions 3 and 5 this influence is visibly lower than in other models.

4.4. Influences on the spatial prediction pattern—empirical method

In this last part, the results for the robustness and sensitivity analysis for the empirical method of the susceptibility map are presented. Outcomes for only one of the trained networks for the subdivision number 9 will be presented. As criterion for the selection of the network, the maximum accuracy calculated on the test set was used.

In order to estimate the robustness of the susceptibility model, a susceptibility map was produced by inserting in each casual factor map the same error model previously used for the calculation of the robustness index. Then the prediction rate curve was calculated for each perturbed model and compared with the original one. As shown in Fig. 13, the prediction capability of the susceptibility map does not change, confirming the robustness of the model already detected when analysing the robustness indexes.

To quantify in a visual and qualitative way the importance of each variable to the final susceptibility map, several susceptibility maps were produced by substituting each variable at the time with random numbers. For each of the obtained maps we calculated the prediction rate curve to detect possible decreases of performance in case the selected variable is substituted with random values. From a qualitative point of view, the variable with less influence on the final susceptibility map is "cosine of aspect". The cross-validation rate curve does not show significant changes, as shown in Fig. 14a, and the pattern in the susceptibility map has not been modified, as shown in Fig. 15b. On the opposite, the variable "PGA" influences the final susceptibility map considerably, since the pattern of the susceptibility model cannot be recognised anymore (Fig. 15c) and the prediction capability visibly drops down Fig. 14b. Other variables (e.g., drainage density, distant to faults, and rainfall) showed similar behaviour as PGA.

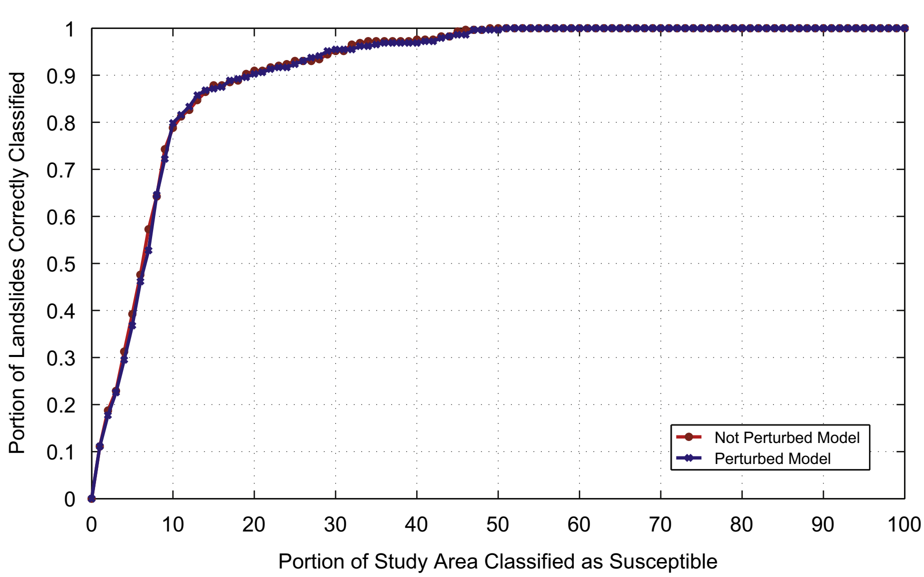

Fig. 13. Prediction rate curve for original and perturbed model (subdivision of the database number 9).

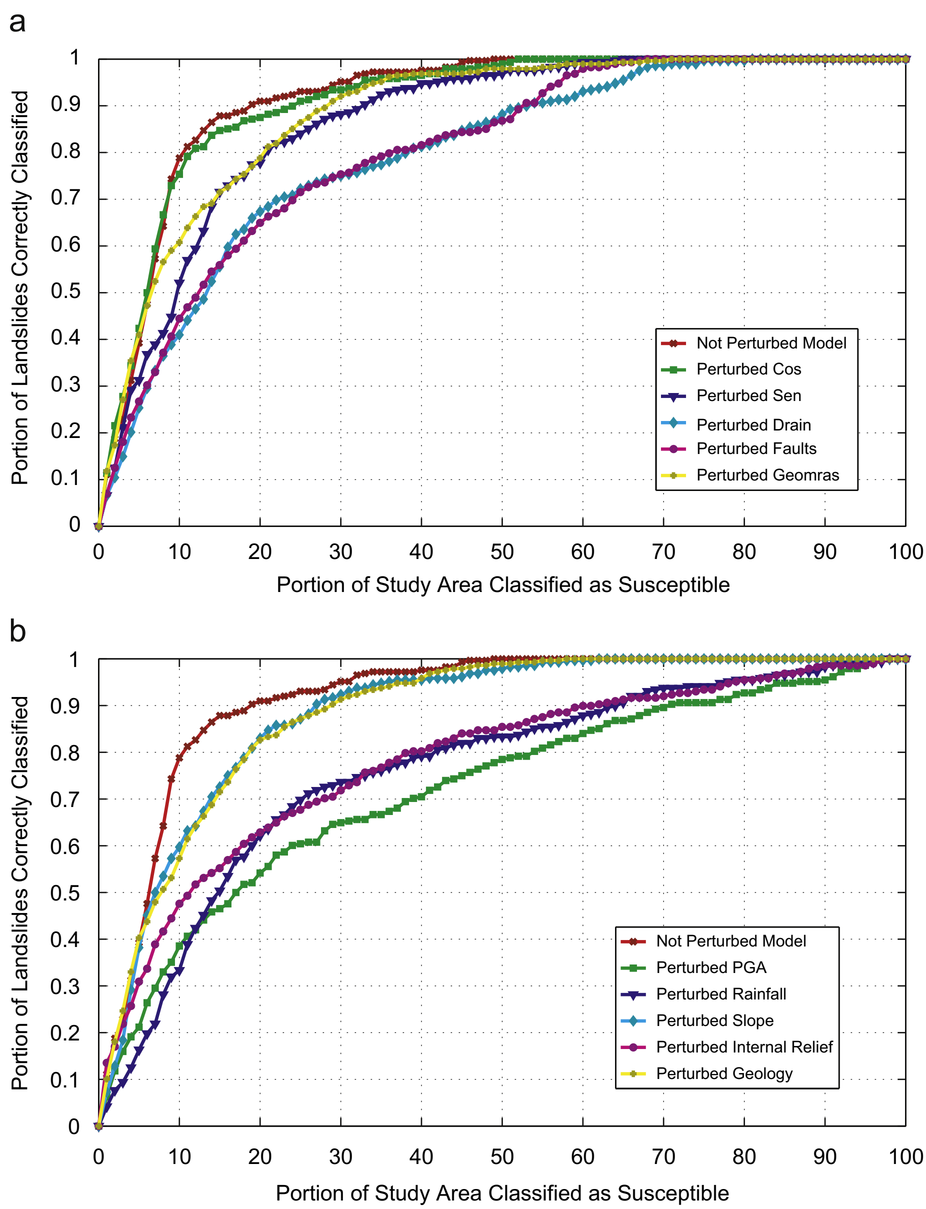

Fig. 14. Prediction rate curve for models in which one variable was left out (subdivision of the database number 9).

5. Conclusions

Although an enormous amount of papers on susceptibility models were published in the last decades, only few of them have dealt with the evaluation of prediction capability, robustness, and sensitivity. Considering the implications that a landslide susceptibility map can have on disaster reduction and mitigation, the lack of well-evaluated models has as consequence the possible production of less correct maps and misinterpretation of the results. As also pointed out in Guzzetti et al. (2006), few publications can be found where the dataset is previously studied by a so-called "exploratory data analysis" and/or where the evaluation of the models is done taking into account several aspects of their quality and not only the prediction capability. Besides that, when the method chosen to model landslide susceptibility is non-linear, as is the case in ANNs, the problem of quality evaluation is made even more complicated due to the complexity of the model and of the lack of already implemented tools to evaluate, for example, the impact of each single variable on the model.

In this paper, we proposed a framework to overcome the majority of the problems related to the evaluation of landslide susceptibility models when the analysis is made by means of ANNs. It was tested in Guantánamo province (Cuba), which is affected by several types of landslides. ANNs were used to assess susceptibility to shallow slides, since the occurrence of this type of mass movement in Guantánamo is complex, and other methods have failed to produce accurate results (Castellanos Abella and van Westen, 2008).

First of all, a 10-fold cross-validation was chosen to evaluate the results of the modelling phase. This technique, as other similar ones based on repetition of the sampling (i.e., leave-one-out cross-validation, and bootstrapping), allows estimating the reliability of the classification, that a simple hold-out validation does not guarantee.

In order to estimate the prediction capability we used 2 methods already widely used and tested: one to estimate the prediction rate on subsets, and the other one to estimate the prediction rate of the final susceptibility map. The robustness was

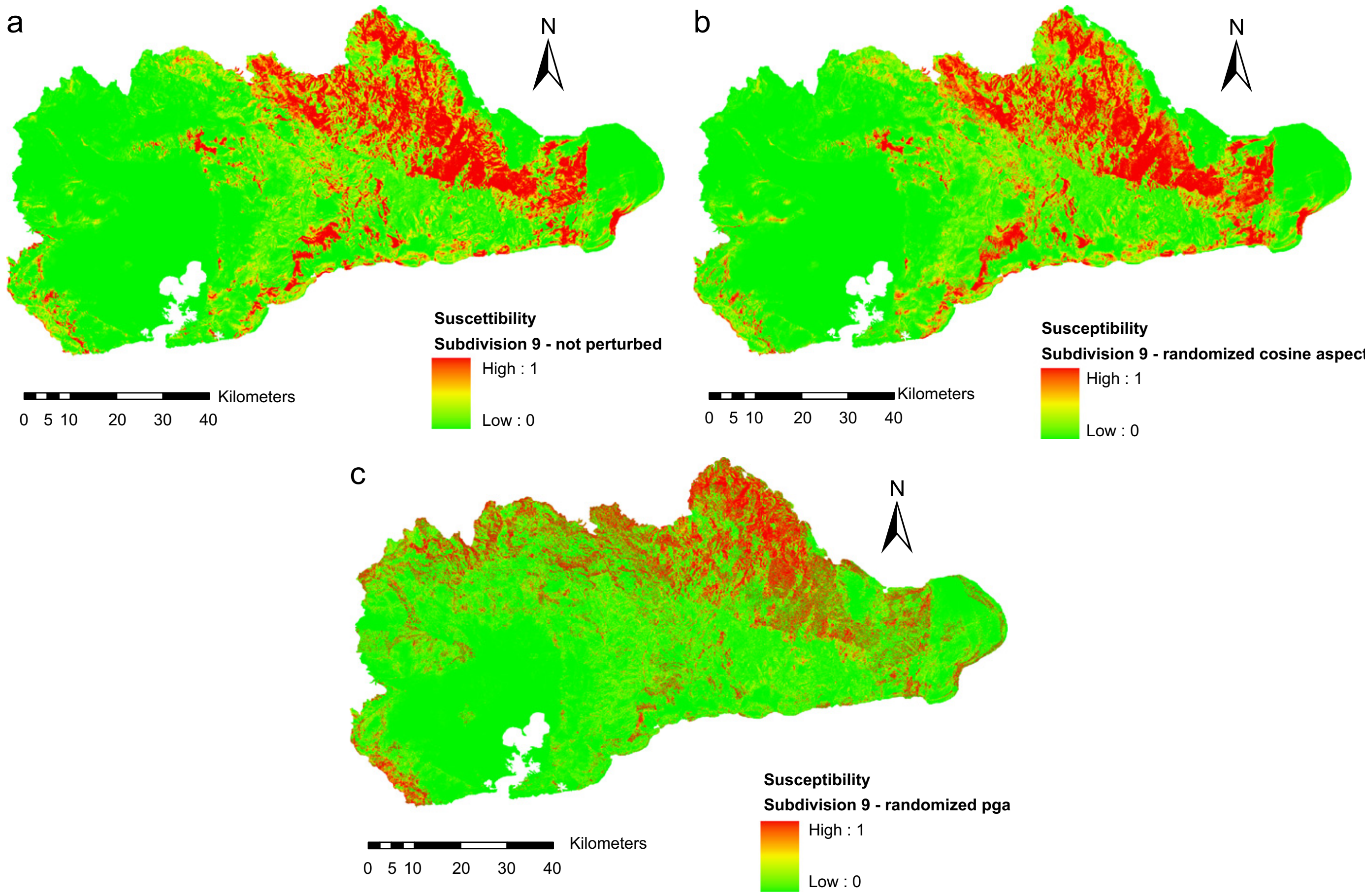

Fig. 15. Susceptibility models for subdivision 9 (a), with the variable "cosine aspect" substituted by random numbers (b), and with the variable "PGA" substituted by random numbers (c).

estimated by using an index approach, previously used by one of the authors. This approach allows estimating the robustness of the model to perturbation (e.g., errors) in each single variable and in all variables at the same time. For the sensitivity analysis a graphical method was tested, that was never previously used in landslide susceptibility modelling. This approach has the advantage that it is easily readable, since it is a graphical method, and allows to detect at the same time: linear and non-linear effects, strength effects, and variable interaction.

The proposed methods for robustness and sensitivity analysis can be used not only in the final evaluation of the models, but also in the model building, as synergic approach for model selection and evaluation. For example, the robustness index and the plots of the sensitivity should help in the selection and removal of casual factors that contain only noise or do not significantly contribute to the model.

We can conclude that:

- By means of ANNs it is possible to obtain models with high prediction capability and high robustness, and that an exploration of the effect of the individual variables is possible, even if they are considered black-box models.
- The results can still have quite a high variance, but the use of 10-fold cross-validation, or similar methods, can help to distinguish in which case the variance is acceptable. With only one run of the model, resulting in one landslide susceptibility map, this would not be possible. We recommend to prepare the susceptibility map taking the results obtained by repetition of the sampling into account. As proposed in this contribution, the calculation of the mean value and of the standard deviation of the 10 networks resulting from the 10-fold cross-validation is a valid method to estimate the susceptibility level and its variability.
- Prediction capability, robustness, and sensitivity are different aspects of model quality, so they should all be taken into account in the modelling phase. Often they provide mutually important results for the understanding of the models. For example the high variance in the prediction capability can be partly explained by some differences in the effect that the causal factors have in the 10 obtained different models.

Moreover, we proposed methods non-specific for ANNs and non-parametric, so they can be used when other statistical or data-driven techniques are chosen to assess landslide susceptibility. This allows comparing different susceptibility models obtained with several mathematical frameworks, taking into consideration 3 components of the quality of model: prediction, robustness, and sensitivity.

Acknowledgements

The authors would like to thank the national organizations in Cuba that provided the necessary data for this study: the Institute of Meteorology, Institute of Geology and Palaeontology and the National Centre of Seismological Research.

References

Alippi, C., Sana, D., Scotti, F., 2004. A training-time analysis of robustness in feed-forward neural networks. In: Proceedings of the International Joint Conference on Neural Networks IJCNN04, Budapest, Hungary, pp. 2853–2858.

Altman, D.G, Bland, J.M., 1994. Diagnostic tests. 1: sensitivity and specificity. British Medical Journal 308 (6943), 1552.

Baeza, C., Corominas, J., 2001. Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. Earth Surface Processes and Landforms 26, 1251–1263.

Bernknopf, R.L., Campbell, R.H., Brookshire, D.S., Shapiro, C.D., 1988. A probabilistic approach to landslide hazard mapping in Cincinnati, Ohio, with application for economic evaluation. Bulletin of the Association of Engineering Geologists XXV (1), 39–56.

Binaghi, E., 1998. Slope instability zonation: a comparison between certainty factor and fuzzy Dempster–Shafer approaches. Natural Hazards 17 (1), 77–97.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Oxford 504 pp.

Cannon, A.J., McKendry, I.G., 2002. A graphical sensitivity analysis for statistical climate models: application to Indian monsoon rainfall prediction by artificial neural network and multiple linear regression models. International Journal of Climatology 22, 1687–1708.

Caruana, R., Lawrence, S., Giles, C.L., 2000. Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping. In: Proceedings of the Neural Information Processing Systems 2000, Denver, Colorado, USA, pp. 402–408.

Castellanos Abella, E.A., 2008. Provincial landslide risk assessment. In: Castellanos Abella, E.A. (Ed.), Multi-Scale Landslide Risk Assessment in Cuba, Utrecht, Utrecht University, ITC Dissertation, vol. 154, pp. 101–152, ISBN: 978-90-6164-268-8.

Castellanos Abella, E.A., van Westen, C.J., 2008. Spatial landslide risk assessment in Guantánamo province, Cuba. In: Proceedings of the 10th International Symposium on Landslides and Engineered Slopes, Xi'an, China, 6 pp.

Chi, K.-H., Park, N.-W., Chung, C.-J., 2002. Fuzzy logic integration for landslide hazard mapping using spatial data from Boeun, Korea. In: Proceedings of the Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada, 6 pp.

Chung, C.F., Fabbri, A.G., 1999. Probabilistic prediction models for landslide hazard mapping. Photogrammetric Engineering and Remote Sensing 65 (12), 1389–1399.

Chung, C.F., Fabbri, A.G., 2003. Validation of spatial prediction models for landslide hazard mapping. Natural Hazard 30 (3), 451–472.

Chung, C.-J., Fabbri, A.G., van Westen, C., 1995. Multivariate regression analysis for landslide hazard zonation. In: Carrara, A., Guzzetti, F. (Eds.), Geographical Information Systems in Assessing Natural Hazards. Kluwer Academic Publishers, Netherlands, pp. 107–133.

Coe, J.A., Michael, J.A., Crovelli, R.A., Savage, W.Z., Laprade, W.T., Nashem, W.D., 2004. Probabilistic assessment of precipitation-triggered landslide using historical records of landslide occurrence, Seattle, Washington. Environmental and Engineering Geoscience 10 (2), 103–122. doi:10.2113/10.2.103.

Davis, T.J., Keller, C.P., 1997. Modelling uncertainty in natural resource analysis using fuzzy sets and Monte Carlo simulation: slope stability prediction. International Journal of Geographical Information Science 11 (5), 409–434.

Duda, R.O., Hart, P.E., Stork, D.G., 2000. Pattern Classification. Wiley-Interscience, New York 654 pp.

Ercanoglu, M., 2005. Landslide susceptibility assessment of SE Bartin (West Black Sea region, Turkey) by artificial neural networks. Natural Hazards and Earth System Sciences 5 (6), 979–992.

Ercanoglu, M., Gokceoglu, C., 2004. Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). Engineering Geology 75 (3–4), 229–250.

Ercanoglu, M., Gokceoglu, C., van Asch, T.W.J, 2004. Landslide susceptibility zoning north of Yenice (NW Turkey) by multivariate statistical techniques. Natural Hazards 32 (1), 1–23.

Ermini, L., Catani, F., Casagli, N., 2005. Artificial neural networks applied to landslide susceptibility assessment. Geomorphology 66 (1–4), 327–343.

Gomez, H., Kavzoglu, T., 2005. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. Engineering Geology 78 (1–2), 11–27.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Galli, M, 2006. Estimating the quality of landslide susceptibility models. Geomphology 81, 166–184.

Hagan, M.T., Menhaj, M.B, 1994. Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks 5 (6), 989–993.

Haykin, S. (Ed.), 1999. Neural Networks. A Comprehensive Foundation, second ed. Prentice-Hall, New Jersey, 842 pp.

Hornik, K.M., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2 (5), 359–366.

Kanungo, D.P., Arora, M.K., Sarkar, S., Gupta, R.P., 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Engineering Geology 85 (3–4), 347–366.

Komac, M., 2006. A landslide susceptibility model using the analytical hierarchy process method and multivariate statistics in perialpine Slovenia. Geomorphology 74 (1–4), 17–28.

Lee, S., Choi, J., Min, K., 2004. Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea. International Journal of Remote Sensing 25 (11), 2037–2052.

Lee, S., Evangelista, D.G., 2006. Earthquake-induced landslide-susceptibility mapping using an artificial neural network. Natural Hazards and Earth System Sciences 6 (5), 687–695.

Lee, S., Ryu, J.-H., Lee, M.-J., Won, J.-S., 2003a. Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea. Environmental Geology 44, 820–833.

Lee, S., Ryu, J.-H., Min, K., Won, J.-S., 2003b. Landslide susceptibility analysis using GIS and artificial neural network. Earth Surface Processes and Landforms 28, 1361–1376.

Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters. SIAM Journal of Applied Mathematics 11, 431–441.

Melchiorre, C., Matteucci, M., Remondo, J., 2006. Artificial neural networks and robustness analysis in landslide susceptibility zonation. In: Proceedings of the International Joint Conference on Neural Networks, Vancouver, Canada, pp. 4375–4381.

Melchiorre, C., Matteucci, M., Azzoni, A., Zanchi, A., 2008. Artificial neural networks and cluster analysis in landslide susceptibility zonation. Geomorphology 94 (3–4), 379–400.

Plate, T., Bert, J., Band, P., 2000. Visualizing the function computed by a feed-forward neural network. Neural Computation 12 (6), 1337–1354.

Saboya, J.F., da Gloria Alves, M., Dias Pinto, W, 2006. Assessment of failure susceptibility of soil slopes using fuzzy logic. Engineering Geology 86 (4), 211–224.