LAND-SE: a software for ~~landslide~~ statistically-based landslide susceptibility zonation, Version 1.0

Mauro Rossi and Paola Reichenbach

CNR IRPI, via Madonna Alta 126, 06128 Perugia, Italia

Correspondence to: Mauro Rossi (mauro.rossi@irpi.cnr.it)

Abstract

Landslide susceptibility (LS) assessment provides a~~n~~ relative estimate of ~~the~~ landslide spatial occurrence based on local terrain conditions. ~~LS has been evaluated in many locations around the world since the early '80 distinct modelling approaches diverse combination of variables and different partition of the territory (mapping units). Among the different methods, statistical models have been largely used to assess LS and several model types have been proposed in the literature. A recent literature review revealed that authors not always present a complete and comprehensive assessment of the LS that includes model performance analysis, prediction skills evaluation and estimation of the errors and uncertainty.~~

A literature review revealed that LS evaluation has been performed in many study areas worldwide using different methods, model types, different partition of the territory (mapping units) and a large variety of geo-environmental data. Among the different methods, statistical models have been largely used to evaluate LS, but the minority of articles presents a complete and comprehensive LS assessment that includes model performance analysis, prediction skills evaluation and estimation of the errors and uncertainty

The aim of this paper is to describe LAND-SE (LANDslide Susceptibility Evaluation), software that performs susceptibility modelling and zonation using statistical models, quantifies the model performances and the associated uncertainty. The software is implemented in R, a free software environment for statistical computing and graphics. This provides users with the possibility to implement and improve the code with additional models, evaluation~~s~~ tools or output types. The paper describes the software structure, explains input

and output, and illustrates specific applications with maps and graphs. The LAND-SE script is delivered with a basic user guide and three example datasets.

Keywords: Landslides, susceptibility, statistical models, zonation, R

1 Introduction

Landslide susceptibility (LS) is the likelihood of a landslide occurring in an area based on local terrain conditions (Brabb, 1984). In mathematical language, LS quantifies the spatial probability of landslides occurrence in a mapping unit, not considering the temporal probability of failure or the magnitude of the expected landslides. Landslide susceptibility has been evaluated in many locations around the world since the early ~~'80~~1980. Authors have evaluated LS using different partitioning of the territory as mapping units, diversified combination of explanatory variables and distinct methods. Methods for the LS evaluation and mapping can be broadly grouped in: geomorphological mapping, analysis of landslide inventories, heuristic or index-based methods, statistically based models and geotechnical or physically based models (Guzzetti et al., 1999). Among the different approaches, the statistical models have been largely used to assess LS. A recent revision of papers on statistical models (Malamud et al., 2014), have shown that more than 95 different model types were proposed in the literature. Malamud and his co-authors grouped them in 20 classes, with the most frequent corresponding to logistic regression, neural networks and data overlay. According to them, the ~~The~~ relevant number of statistical models described in the literature is probably related to the recent increasing number of commercial and open source packages for statistical analysis that can combine and integrate geographical data and/or Open Source GIS (i.e. SAGA GIS, GRASS GIS). The review analysis also revealed that authors not always present a complete and comprehensive assessment of the models performance, ~~and~~ the prediction skills evaluations and the estimation of ~~the~~ errors and uncertainty. Since the large variety of applications of statistical approaches, but the scarcity of model evaluation and quantification of the errors, we have implemented LAND-SE (LANDslide Susceptibility Evaluation), a software developed to prepare landslide susceptibility models and zonation at basin and regional scale, with specific functions focused ~~to~~ on results evaluation and uncertainty estimation. The software is implemented in R, a free software environment for statistical computing and graphics (R Core Team, 2015). This provides users with the possibility to implement and improve the code with additional models, evaluation~~s~~ tools or output types.

The paper describes LAND-SE structure, explains input and output, illustrates them with maps and graphs, some applications and provides a basic user guide. It is out of the scope of the manuscript, the description of the characteristics and results of ~~each~~ statistical models~~,~~

and the advantage/disadvantage of ~~the~~ model evaluation ~~parameters~~ tools and matrixes~~and the analysis of the model results~~. We have introduced a test area only to show and demonstrate possible potential applications and different output of LAND-SE.

The manuscript is structured as follows: section 2 describes the software, its modelling approaches and the main output types; section 3 illustrates the test area to illustrate the range of applications and different outputs of LAND-SE~~and describes some applications~~ and section 4 formalizes some final remarks. The paper is completed by supplementary ~~ancillary~~ materials containing the software code, a user guide and example datasets.

2 Software description

LAND-SE, a software for landslide susceptibility modelling and zonation was implemented and improved with respect to the code proposed by Rossi and co-authors in 2010. The new version is coded in R (R Core Team, 2015) and it is open source. The software holds on the possibility to perform and combine different statistical susceptibility modelling methods, evaluate the results and estimate the associated uncertainty. As compared to the previous version (Rossi et al., 2010), the main improvements are related to: i) the possibility to use different cartographic units (pixel-based vs polygon-based); ii) the capacity to perform different type of validation analyses (spatial, temporal, random); iii) the ability to evaluate the model prediction skills and performances using success and prediction rate curves (Chung and Fabbri, 1999; 2003); iv) the possibility to provide results in standard geographical formats (shapefiles, geotiff); v) an optimization and stabilization of the modelling algorithms; vi) the possibility to use additional computational parameters to tune the calculation procedure, for the analysis of large datasets. This software version presents a relevant computer code restructuring (code refactoring), allowing the implementation of new single statistical approaches (e.g. support vector machines, regression tree based approaches) that can be added as new modules, preserving the basic software structure. The new structure simplifies the maintainability and improvement of the source code.

Figure 1 shows the logical schema of LAND-SE subdivided in the following five functions:

I. Input data ~~Data input~~ preparation;
II. Single susceptibility models and zonation;
III. Combination of single models using a logistic regression approach;
IV. Evaluation of single and combined LS models;
V. Estimation of uncertainty of single and combined LS models.

2.1 Data input preparation

The input data preparation, follows two steps: i) the choice of the cartographic unit and ii) the selection of the criteria for the definition of the training and the validation dataset.

LAND-SE is designed to use different cartographic units, reducible to pixels or to polygon-like subdivisions (e.g. slope units, geomorphological subdivisions, administrative boundaries, etc.). The input data shall be provided in tabular format where each line represents one mapping unit with the associated attributes. Since raster data cannot be used directly as input, a preliminary conversion is required to perform the pixel-based analysis.

The choice of the mapping unit is crucial because it also determines how landslides are sampled to prepare the training and prediction (validation) subsets for the susceptibility modelling. In grid-based susceptibility assessments, several strategies are used to sample landslide pixels, the more frequent are: (1) single pixel selected as the centroid of the entire landslide or the scarp area; (2) all the pixels within the entire landslide body or the scarp area; (3) the main scarp upper edge (MSUE) approach which selects pixels on and around the landslide crown-line; and (4) the seed-cell approach that selects pixels within a buffer polygon around the upper landslide scarp area and sometimes part of the flanks of the accumulation zone (Atkinson et al.,1998; Atkinson and Massari, 2011; Goetz et al., 2015; Heckmann et al., 2014; Hussin et al., 2016; Regmi et al., 2014; Van Den Eeckhaut et al., 2006). The analysis of model sensitivity to different landslide mapping strategies and the significance of different variables combinations can be performed using LAND-SE preparing different input files. Given the numerous possibility of variations required to set this type of evaluation, we decided not to include such functionalities in the current LAND-SE release, but we designed and implemented a command line interface (see §S5 of the LAND-SE User Guide V 1.0) to make this analysis possible using external procedures.

To identify and separate the training and the validation dataset, different criteria can be adopted. Temporal, spatial or random subdivisions can be selected guiding the type of validation analysis. When the temporal validation is selected, secondary information not used in the model training must be provided for the area under analysis. Adopting a temporal subdivision approach, the training and the validation set are composed by the same mapping units and the analysis is performed using the same explanatory variables but different grouping variable (e. g. a different landslide inventory map, often more recent than what is used during the training phase). Differently, in the spatial and random approach, the training and the validation dataset contain different mapping units, characterized by different grouping and explanatory variables. The main difference between the spatial and the random validation is the method chosen to separate the training and the validation dataset: in the first case, the datasets are spatially different (the two areas can be contiguous or not), in the second the subdivision is performed using a random selection. For the pixel-based approach, the definition of the training and the validation dataset can follow the same criteria, but in the literature, the subdivision is commonly performed using a random selection (Van Den Eeckhaut et al., 2010; Felicísimo et al., 2013; Petschko et al., 2014).

2.2 Single susceptibility models estimation (single susceptibility maps)

LAND-SE uses different supervised multivariate statistical models to evaluate the landslide spatial probability, identifying and quantifying the relation between dependent and independent variables. According to previous studies (Carrara et al., 1991; Rossi et al., 2010; Guzzetti et al., 2006, 2012), dependent variable (or grouping variable) is computed as the absence/presence of landslides in the mapping units and is usually derived from a landslide inventory. The independent variables (explanatory variables) are obtained from available thematic information (morphometry, land cover/use, lithology, etc.). Each model is executed in two phases: a the training phase, where the model reconstructs the relationships between the dependent and the independent variables, and a validation phase, where these relationships are verified in different conditions. LAND-SE calculates landslide susceptibility with ~~the following~~four single models (Rossi et al., 2010): i) linear discriminant analysis (LDA) (Fisher, 1936; Brown, 1998; Venables and Ripley, 2002), ii) quadratic discriminant analysis (QDA) (Venables and Ripley, 2002), iii) logistic regression (LR) (Cox, 1958; Brown, 1998; Venables and Ripley, 2002), and iv) neural network (NN) modelling (Ripley, 1996; Venables and Ripley, 2002). The logistic regression model was significantly improved with respect to Rossi et al. (2010), substituting the previous code based on the “Zelig” package (Owen et al., 2013), with a more stable and performing code based on the “glm” function, included in the well tested base R implementation (R Core Team, 2015).

2.3 Combined model using a logistic regression approach (combined susceptibility maps)

~~Similarly to the previous version,~~ LAND-SE uses a combination model (CM) based on a logistic regression approach, where the grouping variable is the presence or absence of landslides in the mapping units, and the explanatory variables are the forecasts of the selected single susceptibility models (Rossi et al., 2010). Similarly, to the single logistic regression model, the original code based on the “Zelig” package was substituted with the “glm” function. LAND-SE allows to enable or not, the execution of the combined model selecting different combinations of single models.

2.4 Susceptibility model evaluation

In the training phase, LAND-SE reconstructs the relationships between dependent and independent variables and evaluates the prediction skills of single and combined models (i.e. the capability to predict the original data). In the validation phase, LAND-SE verifies the results in different conditions and evaluates the models capability to predict independent data. Model~~s~~ outputs of both phases are evaluated using the same tools and in particular the following statistical metrics and indices:

- The dependence among explanatory variables (Belsley, 1991; Hendrickx, 2012);
- Contingency tables (i.e. confusion matrixes) (Jollifee and Stephenson, 2003);
- Contingency plots or fourfold plots summarizing the mapping units correctly and incorrectly classified by the models (Jollifee and Stephenson, 2003);
- Error maps showing the geographical distribution of the mapping units correctly and incorrectly classified by the models (Rossi et al., 2010);
- Plots showing receiver operating characteristic (ROC) curves (Green and Swets, 1966; Mason and Graham, 2002; Fawcett, 2006) and the associated Area Under Curve (AUC) statistics;
- Evaluation plots showing the variation of the sensitivity (“hit rate”), the specificity (1-false positive rate), and of the Cohen's kappa index (Cohen,1960);
- Success and prediction rate curves (Chung and Fabbri, 1999; 2003)

The description and discussion of the characteristics, advantage/drawbacks of these statistical metrics/indices are out of the scope of the manuscript and they will not be described in detail.

2.5 Uncertainty evaluation (single and combined susceptibility zonations)

For each single and combined model, LAND-SE evaluates and quantifies the uncertainty adopting a "bootstrapping" ~~re-sampling technique~~approach. Bootstrapping is a resampling technique for estimating the distributions of statistics based on independent observation. Bootstrapping can refer to any test or metric that relies on random sampling with replacement (Efron, 1979; Davison and Hinkley, 2006). The technique has been largely used to estimate errors and uncertainties associated to classification models (among the others, Kuhn and Kjell, 2013). In the training phase, a user-specified number of runs are performed varying the selected dataset. Descriptive statistics for the probability (susceptibility) estimates, including the mean (μ) and the standard deviation (σ), are obtained from an ensemble of model runs (i.e. a user-defined number of LAND-SE simulations are executed to obtain the two descriptive statistics). Such information is portrayed in plots showing estimates for the model uncertainty in each mapping unit and in maps showing the geographical distribution of the uncertainty (Guzzetti et al., 2006; Rossi et al., 2010). To model the uncertainty associated to each LS zonation, the mean and the standard deviation are fitted using a parabolic function (Figure 3D). Such function is used to estimate the uncertainty in the validation phase. The map showing the geographical distribution of the uncertainty can provide additional and relevant information for the use of LS zonation in environmental planning studies. A proper interpretation of the map may provide for each mapping unit a proxy of a degree of confidence associated to the LS estimate.

The sampling procedure implemented in LAND-SE is only focused to the estimation of the uncertainty associated to the susceptibility zonation. However, the software also outputs estimates of the performance variability in the training and validation phases providing confidence levels in the ROC plots (NCAR, 2014) and in the fourfold or contingency plots (Meyer et al., 2015). In addition, the execution of analyses that investigate sensitivity or variability of model results when varying inputs (e.g. using sampling procedures), is facilitated by the LAND-SE command line interface, that makes these analyses possible using external procedures.

2.6 ~~SW~~ Software output formats

LAND-SE can be executed in two different modes: the *standard* that provides textual and graphical results stored respectively in .txt and .pdf, and the *geomode* providing also geographical output as shapefiles and GeoTIFF. Some output (i.e the success and prediction rate curves) are produced only in the *geomode* because they require geographical data (shapefile) as additional input. A complete list of the output with a detailed description is provided in the supplementary material (LAND-SE_UserGuide.pdf).

3 LAND-SE applications

To show LAND-SE ~~software~~ functionalities and output types, we use as example the landslide susceptibility modelling and zonation originating from two articles published by Reichenbach and co-authors (2014; 2015)~~LAND-SE was applied in a test area~~. In the area selected as example, ~~Different configurations were selected to~~we perform the following analysis, using different configurations:

- Polygon-based landslide susceptibility zonation;
- Pixel-based landslide susceptibility zonation;
- Landslide susceptibility scenarios zonation.

The applications use different mapping units and distinct schema to select the training and validation dataset. ~~One~~ The last analysis ~~is focused to~~ illustrates an application ~~the use of LAND-SE~~ focused to evaluate the impact of different land use scenarios ~~of land use~~ on landslide susceptibility~~LS~~. ~~LAND-SE~~ This type of analysis and results can be ~~considered~~ relevant information in environmental planning and management.

3.1 Description of the example area and available data

A small area was selected to show applications and output of LAND-SE. The area is located in the eastern portion of the Briga catchment (Figure 2), in the Messina province (Sicily, South Italy). ~~It has~~The elevation ~~values ranging~~ ranges from the sea level to about 500 m and the terrain gradient ~~in the range of~~from 0° ~~–~~to ~~81~~80°. Landslides, including shallow soil slides and debris flows, deep-seated rotational and translational slides, and complex and compound failures (Varnes, 1984), are abundant, and caused primarily by rainfall (Ardizzone et al., 2012; Reichenbach et al., 2014; 2015). On 1 October 2009, the Briga catchment and the surrounding areas were hit by an intense storm (Maugeri and Motta, 2011) that triggered more than 1000 shallow landslides, mainly shallow soil slides and debris flows (Varnes, 1984), caused 37 fatalities, numerous injured people and severe damages in the affected villages and along the transportation network.

After the event, a detailed landslide inventory map at 1:10,000 scale was prepared for the entire Briga catchment (Ardizzone et al., 2012). The inventory was obtained through a combination of field surveys carried out in the period from October to November 2009, and visual interpretation of pre-event and post-event stereoscopic and pseudo-stereoscopic aerial photographs. The inventory map shows the distribution and types of landslides triggered by the 1 October 2009 rainfall event (Figure 2), and the distribution and types of pre-existing landslides. In addition, two maps reporting the land use in different periods were prepared exploiting available aerial photographs and Very High Resolution (VHR) satellite imagery (Reichenbach et al., 2014; 2015). The first map was derived from the analysis of the same black and white aerial photograph used to map pre-event landslides. The second map was obtained from the analysis of two QuickBird satellite images taken the first on 2 September 2006 and the second on 8 October 2009 (Mondini et al., 2011).

In the area, landslide susceptibility zonation ~~were~~ was prepared using two mapping units: pixels and slope-units. The slope-units (SU) are terrain subdivisions bounded by drainage and divide lines (Carrara et al., 1991). SU were outlined using a 5-meter resolution DEM obtained resampling the VH resolution DEM provided by the Italian national Department for Civil Protection and using *r.slopeunits,* a software recently ~~developed *r.slopeunits*~~ written in Python for GRASS GIS ~~module~~ (Marchesini et al., 2012;–Alvioli et al., 2016). The size and the geometrical characteristics of the SU are controlled by modeling parameters defined by the user including the minimum half-basin area (Metz et al., 2011) and the slope aspect variability. In the study area, the procedure identified 238 SU which represent the polygon-based mapping units for the determination of LS. To explain the spatial distribution of landslides (Carrara et al., 1991; 1995), for each slope-unit, we calculated the percentage of the event landslides as dependent (grouping) variable and the following explanatory variables: i) descriptive statistics (range, mean, standard deviation) of elevation and slope; ii) the percentage covered by each land use class; and iii) the percentage covered by old landslides.

For the pixel-based analysis, we used the VH resolution DEM (1m x 1m) that accounts for about 5 million cells. Maps of the elevation, slope, land use and of the presence/absence of old landslides, were used as explanatory variables in the analysis. The presence/absence of event landslides was used as dependent variable (Carrara et al., 1991, 1995; Guzzetti et al., 2006). The data originally in polygon format were first converted in raster and all the data were converted to the tabular format to be suited for LAND-SE (see LAND-SE_UserGuide.pdf for details).

3.2 Polygon-based landslide susceptibility zonation

This example is focused to illustrate landslide susceptibility zonation prepared using the slope-unit as mapping unit. Two spatial criteria were used to define the training and validation dataset, the first based on a random selection and the second on the subdivision of the entire catchment in two contiguous areas (Nord and South).

In the first case, the training set contained 70% of the total slope-units and the validation corresponded to the entire basin. Landslide susceptibility models were trained using a subset of available data and results were applied in validation to the entire study area. Figure 3 shows the main graphical and geographical outputs obtained during the training and the validation phases, including susceptibility, error and uncertainty maps, fourfold (contingency) plot, success and prediction rate curves, ROC plot, evaluation and uncertainty plots. For simplicity, the figure shows only results of the combined model, but outputs for each single model are available and can be exploited for further analysis. In the example, the random selection criteria resulted in similar training and validation performances (Figure 3). This application simulates LS zonation for a large territory, where landslides information is spotted and do not cover the entire study area. In such conditions, training cannot be performed on the entire area and a random selection of the training dataset, within the surveyed area, is a reasonable solution.

In the second case, the SU located in the Northern part of the Briga catchment with respect to the main river, were used as training set and the SU located in the Southern portion as validation set. Figure 4 shows outputs, including susceptibility maps for the combined model, success and prediction rate curves, and ROC plots. As shown in Figure 4, the spatial subdivision resulted in good model skill analysis, but reduced validation performances, underlying a poor spatial transferability (Ruette et al., 2011; Petschko et al., 2014)~~exportability~~ of the model (i.e. poor applicability of the resulting model coefficients to different study areas). This type of application simulates LS zonation for areas where landslide~~s~~ information required to train the model, is available only for a portion of the area. Results obtained in the training phase are then applied to estimate susceptibility to the portion of the territory where landslide data are not available. This application can be useful to evaluate the possibility to use the same model output in different portion of territory or in different areas.

3.3 Pixel-based landslide susceptibility zonation

This example shows a landslide susceptibility zonation prepared using the pixel as mapping unit. A random selection was chosen to prepare the training set and the validation was performed applying results on the entire study area. For the purpose, in the training set all the pixels corresponding to landslides and an equal number of pixels in stable areas were selected. Figure 5 shows the main outputs of the combined model prepared for the entire area during the validation phase, including susceptibility, error and uncertainty maps, fourfold (contingency) plot, prediction rate curve, ROC plot, evaluation and uncertainty plots.

This example simulates a common and widespread susceptibility zonation approach that exploits pixel-based analysis at basin and regional scale. In such conditions, reasonable calculation times ~~with a limited loss of performances~~ can be reached ~~to~~ training the model with a random selected subset and applying results to the entire study area. Dealing with large dataset, we experienced that training the models using reduced samples (randomly selected) affects slightly the susceptibility model results and performances with a minor increase in the model uncertainty. As shown in Figure 5, although the training was performed with a subset of the data, the model performance for the entire study area is adequate and acceptable.

3.4 Landslide susceptibility scenarios zonation

This example illustrates how LAND-SE can be utilized to evaluate the impact of different land-use scenarios on landslide susceptibility zonation (Reichenbach et al. 2014, 2015) comparing the distribution of stable/unstable slope units and the success rate curves. The current, the past and possible future land-use distributions were evaluated on landslide susceptibility classes. Single models (linear discriminant analysis, quadratic discriminant analysis and logistic regression) and a combined model were prepared, exploiting the 2009 event landslides as grouping variable and morphological and land-use classes as explanatory variables.

To evaluate the influence of land use change on landslide susceptibility zonation, results obtained with the 2009 land use map were applied using the 1945 land use distribution. Figure 6 portrays on the left, the combined model prepared using the 2009 land use map, and on the right the zonation obtained applying the results to the 1954 land use cover. Zonation maps obtained with the same models but using the 1954 land use map show a significant reduction in the number of unstable SU. Success rate curves reveal a decrease in the model fitting performance when using the 1954 land use map, due to a reduction of slope units classified as unstable and an increase in stable terrain. In particular, the expansion of bare soil to the detriment of forested areas in the 56 years from 1954 to 2009, determined a general increase in the susceptibility.

Moreover, to estimate the effect of land use distribution, we have designed different scenarios obtained changing the 2009 land use cover using and heuristic and empirical approach. Assuming an increase in the forested areas, we have considered three types of changes computed at the slope unit scale resulting in the following scenarios: i) 75% decrease in the pasture extent (Scenario 1); ii) 75% reduction of both pasture and cultivated areas (Scenario 2); and iii) 75% decrease in bare soil where the slope-unit mean angle was greater than 15° together with 75% decrease in pasture areas (Scenario 3). A fourth scenario was prepared assuming the effect of a forest fire in the south-west part of the area, where we simulated a reduction of the forested cover and an increase in bare soil (Scenario 4). For each scenario, figure 7 shows the CM zonation and the success rate curve measuring the fitting performance of each model.

The qualitative comparisons of the maps and of the success rate curves obtained for the different~~Analyses of the~~ scenarios confirm how land use changes significantly affect the spatial distribution of unstable/stable slopes (Reichenbach et al., 2014). This information can be ~~used~~ applied to evaluate the consequences of land use change on vulnerability and risk. Moreover, the proposed approach can be helpful to analyse the potential effects of land use planning and management on slope instability.

4 Final remarks

A recent review analysis on landslide statistical models revealed a large variety of statistical types, but a significant scarcity of a complete and comprehensive evaluation of the models performance and prediction skills (Malamud et al. 2014). Moreover assessment of the input data quality (Ardizzone et al. 2002), discussion on the scale applicability and the quantification of errors and uncertainty associate to the models are limited. In the recent years there has been an increase number of commercial and open source packages for statistical analysis that integrate geographical data and/or Open Source GIS, but software dedicated to landslide susceptibility zonation using statistical models is not available.

LAND-SE is an open source ~~SW~~software that performs LS modelling, zonation, results evaluation and associated uncertainty estimation using graphs, map and statistical metrics filling the lacks of the large variety of statistical methods already available. LAND-SE is mainly designed to evaluate landslide susceptibility from basin (medium) to regional scale (small to very small scale).The quality and significant of model outputs is highly related to the scale, accuracy and resolution of landslide and environmental input data. In the field of landslide susceptibility zonation, LAND-SE is designed to be properly and productively used by experienced geomorphologists. Experienced practitioners are expected to use the code, with the support of experts in the field of environmental planning and management for a correct and reliable interpretation and exploitation of the results. A proper LAND-SE execution requires: (i) a basic knowledge of R language to run the script; (ii) experience on multivariate statistical models and on their evaluation skills/metrics (ROC plot, contingency table and plots, success/prediction rate curves, etc.); (iii) GIS skills to prepare and handle input data; and (iv) specific expertise for a correct and reliable interpretations of the results. All the modelling types implemented in LAND-SE are basically statistical classification techniques applicable to any multivariate analysis with a binary grouping (dependent or response) variable. This makes the code flexible and appropriate to other scientific fields and usable, with minor customization and tailoring, by user with different expertise.

We think further improvements may include additional models (i.e. forest tree analysis), tools for the input data preparation, tools for the visualization of results available now only in textual format (i.e. test of the collinearity evaluation, number of significant variables). Moreover, the software can be applied and customized to different applications, providing the users with the possibility to implement and improve the code with additional models, evaluations tools or output types. LAND-SE can also be used to prepare models to predict particular types of slope movements (e.g. debris flow source areas, Carrara et al., 2008) or can be customized to evaluate the probability of spatial occurrence of completely diversified natural phenomena.

Acknowledgements

The implementation and improvement of LAND-SE with respect to the version published by Rossi et al. (2010), was supported by the FP7 LAMPRE Project (Landslide Modelling and Tools for vulnerability assessment preparedness and recovery management, EC contract n. 31238).

Code availability and licence

The LAND-SE code is provided as supplementary materials together with:

1. the software user guide (LAND-SE_UserGuide_v1_03mar2016.docx);
2. datasets containing the software script (LAND-SE_v30_20160118.R), the configuration files (LAND-SE_configuration_spatial_data.txt, LAND-SE_configuration.txt) and input files (training.txt, training.shp, validation.txt, validation.shp) relative to three examples applications: (i) polygon-based landslide susceptibility zonation with a random selection of the training dataset and a validation on a larger area; (ii) polygon-based landslide susceptibility zonation with training and validation performed in two different contiguous areas; (iii) pixel-based landslide susceptibility zonation with a random selection of the training dataset and a validation on a larger area.

References

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., and Guzzetti, F.: Automatic delineation of geomorphological slope-units and their optimization for landslide susceptibility modelling, Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-118, in review, 2016

~~Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Fiorucci, F., Ardizzone, F., and Guzzetti, F.: Automatic delineation of geomorphological slope-units and their optimization for a selected landslide susceptibility model, Environmental Modelling & Software (under revision), 2016~~.

Ardizzone, F., Basile, G., Cardinali, M., Casagli, N., Del Conte, S., Del Ventisette, C., Fiorucci, F., Garfagnoli, F., Gigli, G., Guzzetti, F., Iovine, G., Mondini, A. C., Moretti, S., Panebianco, M., Raspini, F., Reichenbach, P., Rossi, M., Tanteri, L., and Terranova, O.: Landslide inventory map for the Briga and the Giampilieri catchments, NE Sicily, Italy, Journal of Maps, 8:2, 176-180, 2012.

Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., and Reichenbach P.: Impact of mapping errors on the reliability of landslide hazard maps, Natural Hazards and Earth System Sciences, 2:1-2, 3-14, 2002.

Atkinson, P., Jiskoot, H., Massari, R., and Murray, T.: Generalized linear modelling in geomorphology, Earth Surf. Process. Landf., 23(13), 1185–1195, 1998~~.~~

Atkinson, P. M., and Massari, R.: Autologistic modelling of susceptibility to landsliding in the central Apennines, Italy. Geomorphology, Accepted Manuscript, doi: 10.1016/j.geomorph.2011.02.001, 2011.

Belsley, D.A.: Conditioning diagnostics, collinearity and weak data in regression, John Wiley & Sons, New York, 1991.

Brabb, E. E.: Innovative approaches to landslide hazard mapping, Proceedings 4th International Symposium on Landslides, Toronto, 1, 307-324, 1984.

Brown, C.E.: Applied Multivariate Statistics in Geohydrology and Related Sciences, Springer-Verlag, Berlin, 248 pp., 1998

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., and Reichenbach, P.: GIS techniques and statistical models in evaluating landslide hazard, Earth Surface Processes Landform, 16:5, 427-445, 1991.

Carrara, A., Crosta, G., and Frattini, P.: Comparing models of debris-flow susceptibility in the alpine environment, Geomorphology, 94:3, 353-378, 2008.

Carrara, A., Cardinali, M., Guzzetti, F., and Reichenbach, P.: GIS technology in mapping landslide hazard, Carrara, A., Guzzetti, F. (eds.), Geographical Information Systems in Assessing Natural Hazards. Kluwer Academic Publisher, Dordrecht, The Netherlands, 135-175, 1995.

Chung, C.–J.F., and Fabbri, A.G.: Validation of spatial prediction models for landslide hazard mapping, Natural Hazards, 30:3, 451-472, 2003.

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

Cohen, J.: A coefficient of agreement for nominal scales, Educational and Psychological Measurement, 20, 37-46, 1960.

Cox, D.R.: The regression analysis of binary sequences. Journal of the Royal Statistical Society, Series B, Methodological 20, 215–242, 1958.

Davison, A. C., and Hinkley, D.: Bootstrap methods and their applications, 8th ed., Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, ISBN-13: 9780521574716, 2006.

Efron, B.: Bootstrap methods: another look at the jack knife, Annals Statistics 7:1, 1-26, 1979.

Fawcett T.: An introduction to ROC analysis, Pattern Recognition Letters, 27:8, 861-874, 2006.

Felicísimo, Á., Cuartero, A., Remondo, J., and Quirós, E.: Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study, Landslides, 10:2, 175-189, 2013

Fisher, R.A.: The use of multiple measurements in taxonomic problems, Annales Eugenics 7, 179–188, 1936.

Goetz, J. N., Brenning, A., Petschko, H. and Leopold, P.: Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling, Comput. Geosci., 81, 1–11, doi:10.1016/j.cageo.2015.04.007, 2015.

Green, D. M., and Swets, J. M.: Signal detection theory and psychophysics, John Wiley and Sons, New York. ISBN: 0-471-32420-5, 1966.

Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P.: Landslide hazard evaluation: a review of current techniques and their application in a multi–scale study, Central Italy, Geomorphology, 31: 181-216, 1999.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., and Galli, M.: Estimating the quality of landslide susceptibility models, Geomorphology, 81:1-2, 166-184, 2006.

Jollifee, I. T., and Stephenson, D. B.: Forecast verification, A Practitioner's Guide in Atmospheric Science. John Wiley, Chichester, 240 pp, 2003.

Kuhn, M., and Kjell, J.: Applied predictive modeling. New York: Springer, 2013.

Hendrickx, J.: perturb: Tools for evaluating collinearity. R package version 2.05. http://CRAN.R-project.org/package=perturb, 2012.

Heckmann, T., Gegg, K., Gegg, A. and Becht, M.: Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows, Nat. Hazards Earth Syst. Sci., 14(2), 259–278, doi:10.5194/nhess-14-259-2014, 2014.

Hussin Y. H., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., and Bălteanu, D. (2016) Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. Geomorphology 253, 508–523. http://dx.doi.org/10.1016/j.geomorph.2015.10.030.

Malamud, B., Mihir, M., Reichenbach, P., and Rossi, M.: D6.3-Report on standards for landslide susceptibility modelling and terrain zonations, LAMPRE FP7 Project deliverables, http://www.lampre-project.eu, 2014.

Marchesini, I., Alvioli, M., Rossi, M., Santangelo, M., Cardinali, M., Reichenbach, P., Ardizzone, F., Fiorucci, F., Balducci, V., Mondini, A.C., and Guzzetti, F.: WPS tools to support geological and geomorphological mapping, OGRS 2012, Open Source Geospatial Research & Education Symposium, October 24-26 2012, Yverdon-les-Bains, Switzerland, http://ogrs2012.org/index.php/ogrs2012/ogrs2012/paper/view/34, 2012.

Mason, S. J., and Graham, N. E.: Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation, Quarterly Journal of the Royal Meteorological Society, 128, 2145-2166, 2002.

Maugeri, M., and Motta, E.: Effects of heavy rainfalls on slope behavior: the October 1, 2009 disaster of Messina (Italy), Iai, S. (ed.) Geotechnics and Earthquake Geotechnics Towards Global Sustainability, Geotechnical, Geological and Earthquake Engineering 15, Springer Science + Business Media, 2011.

Metz, M., Mitasova, H., and Harmon, R.: Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search, Hydrology Earth System Science, 15, 667-678, 2011.

Meyer, D., Zeileis, A., Hornik, K.: vcd: Visualizing Categorical Data, R package version 1.4-0, 2015.

Mondini, A. C., Guzzetti, F., Reichenbach, P., Rossi, M., Cardinali, M., and Ardizzone, F.: Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images, Remote Sensing of Environment, 115:7, 1743-1757, 2011.

NCAR - Research Applications Laboratory: verification: Weather Forecast Verification Utilities, R package version 1.41. http://CRAN.R-project.org/package=verification, 2014.

Owen, M., Imai, K., King, G., and Lau, O.: Zelig everyone's statistical software, R package version 4.2-1, http://CRAN.R-project.org/package Zelig, 2013.

Petschko, H., Brenning, A., Bell, R., Goetz, J., and Glade, T.: Assessing the quality of landslide susceptibility maps – case study Lower Austria, Natural Hazards Earth System Science, 14:1, 95-118, 2014

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2015.

Regmi, N. R., Giardino, J. R., McDonald, E. V. and Vitek, J. D.: A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA, Landslides, 11(2), 247–262, doi:10.1007/s10346-012-0380-2, 2014.

Reichenbach, P., Busca, C., Mondini, A. C., and Rossi, M.: Land use change scenarios and landslide susceptibility zonation: the Briga catchment test area (Messina, Italy), Engineering Geology for Society and Territory, 1, 557-561, Springer International Publishing, 2015.

Reichenbach, P., Busca, C., Mondini, A. C., and Rossi, M.: The influence of land use change on landslide susceptibility zonation: the Briga catchment test site (Messina, Italy), Environmental Management, 54:1372-1384, 2014.

Ripley, B.D.: Pattern Recognition and Neural Networks. InCambridge University Press, ISBN: 0-521-46086 7, pp. 416, 1996.

Rossi M., Guzzetti F., Reichenbach P., Mondini A. C., and Peruccacci S.: Optimal landslide susceptibility zonation based on multiple forecasts, Geomorphology, 114:3, 129-142, 2010.

Van Den Eeckhaut, M., Marre, A., and Poesen, J.: Comparison of two landslide susceptibility assessments in the Champagne–Ardenne region (France), Geomorphology, 115:1–2, 141-155, 2010

Varnes D. J.: IAEG Commission on Landslides: Landslide hazard zonation: a review of principles and practice, 1984.

Venables, W.N., Ripley, B.D.: Modern Applied Statistics with S, Fourth edition, Springer, Berlin, ISBN: 0-387-95457-0, pp. 495, 2002.

von Ruette, J., Papritz, A., Lehmann, P., Rickli, C. and Or, D.: Spatial statistical modeling of shallow landslides - Validating predictions for different landslide inventories and rainfall events, Geomorphology, 133: 1–2, 11-22, 2011.

Figures

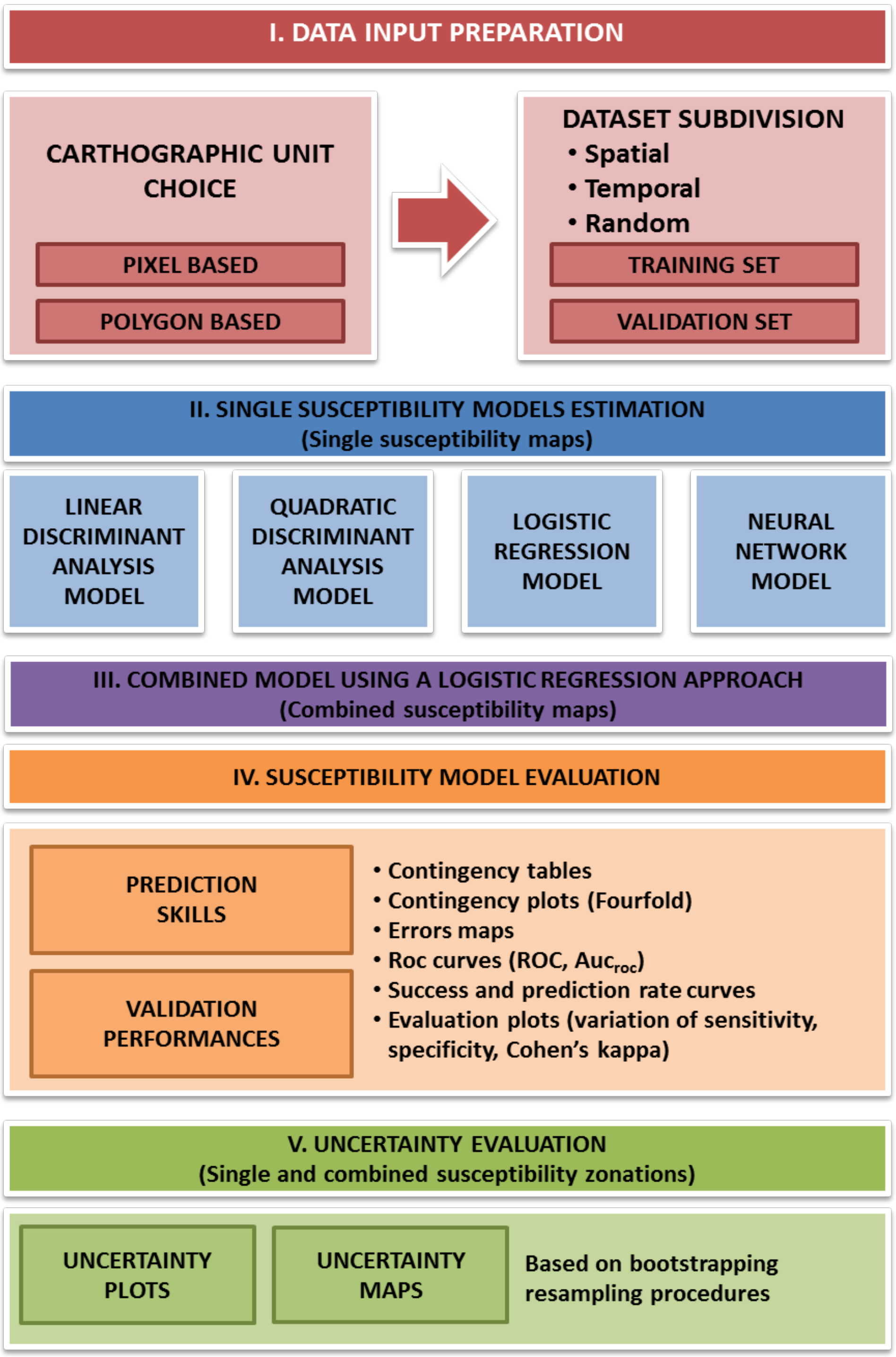
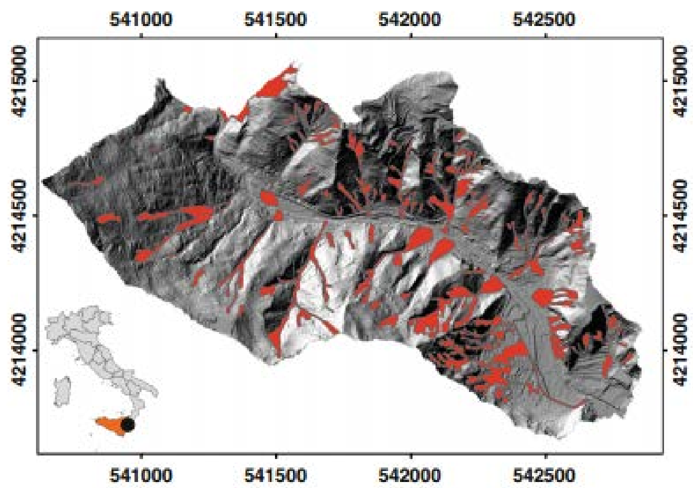


Figure 1. Logical schema of the LAND-SE software for landslide susceptibility modelling and zonation.



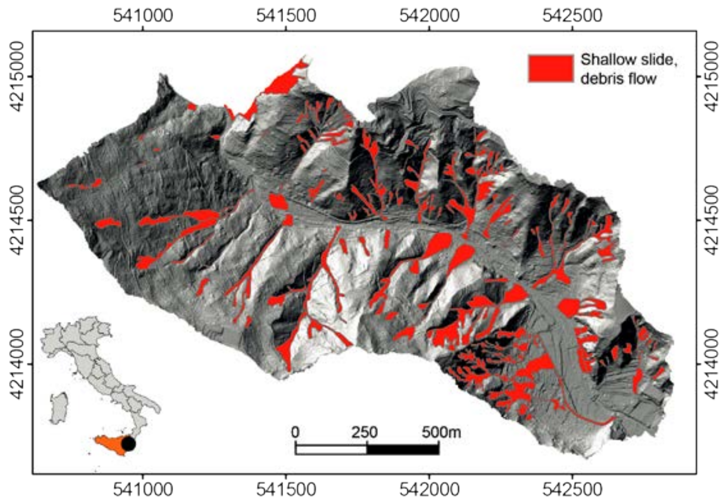


Figure 2. Shaded relief of the study area located in the Briga catchment, along the Ionian coast of Sicily (Italy). Red polygons show landslides triggered by the October 1, 2009 rainfall event.

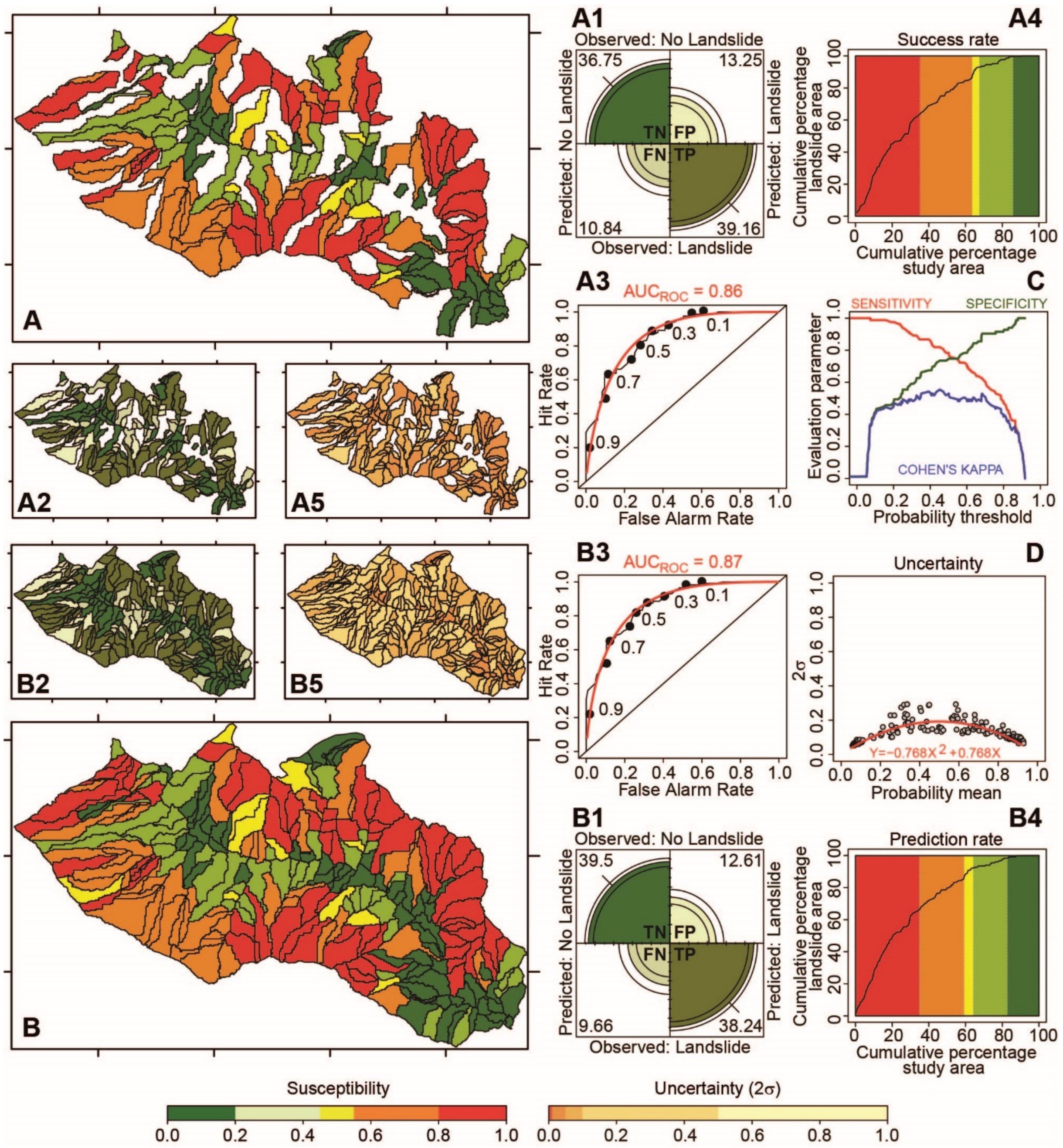


Figure 3. Landslide susceptibility maps (CM) for the training dataset (A) and the validation dataset (B) classified in five unequally spaced classes (see legend). (A1, B1) fourfold plots summarizing the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN); (A2, B2) maps of the distribution of the four categories of slope units reported in the fourfold plots; (A3, B3) ROC plots; (A4, B4) success and prediction rate curves; (C) variation in the model sensitivity, specificity, and Cohen's kappa index; (D) plot showing measures of the model error (2σ) vs. the mean probability (μ), for each slope unit, (black circle); (A5, B5) maps of the geographical distribution of the model error. Maps coordinates and scale bar are shown in Figure 2.

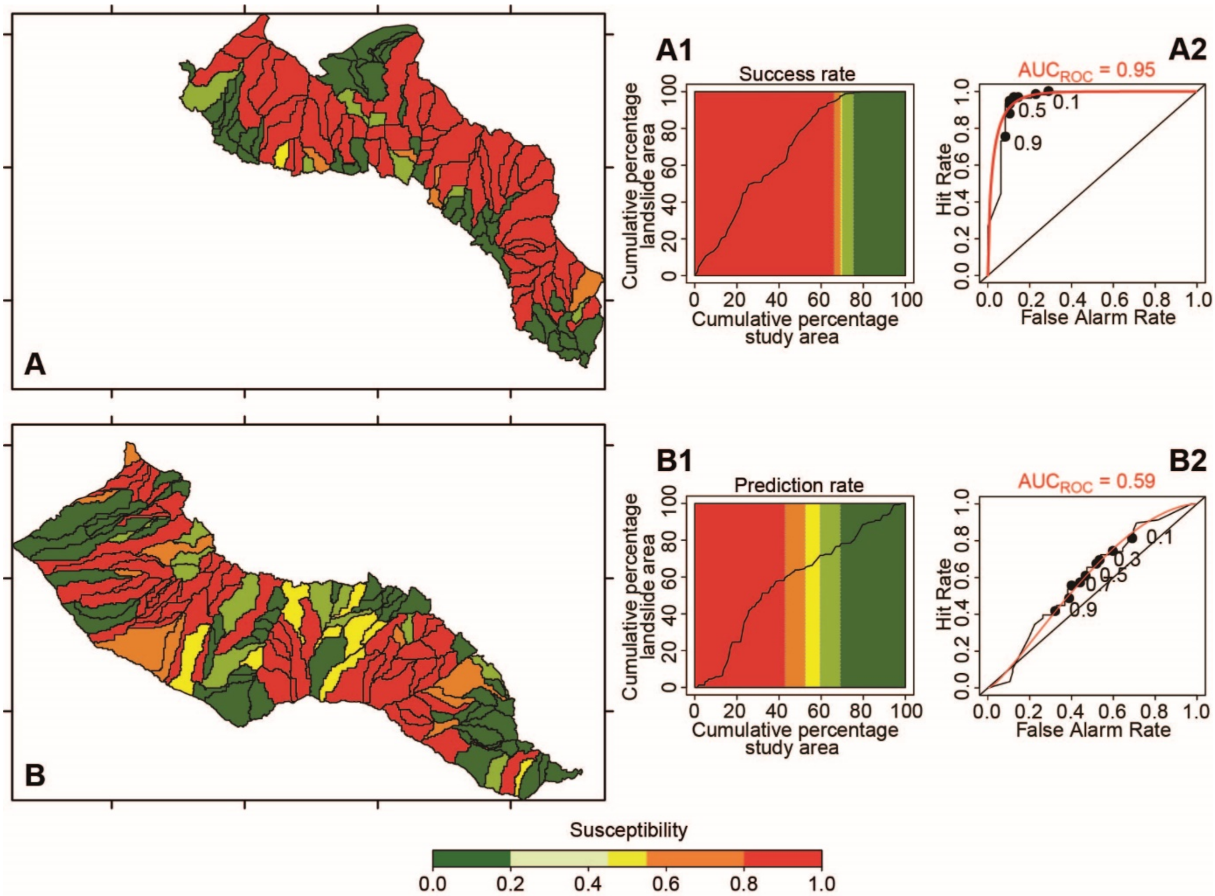


Figure 4. Landslide susceptibility maps (CM) for the training dataset (A: Northern part) and the validation dataset (B: Southern part) of the test area, classified in five unequally spaced classes (see legend). (A1, B1) success and prediction rate curves; (A2, B2) ROC plots. Maps coordinates and scale bar are shown in Figure 2.

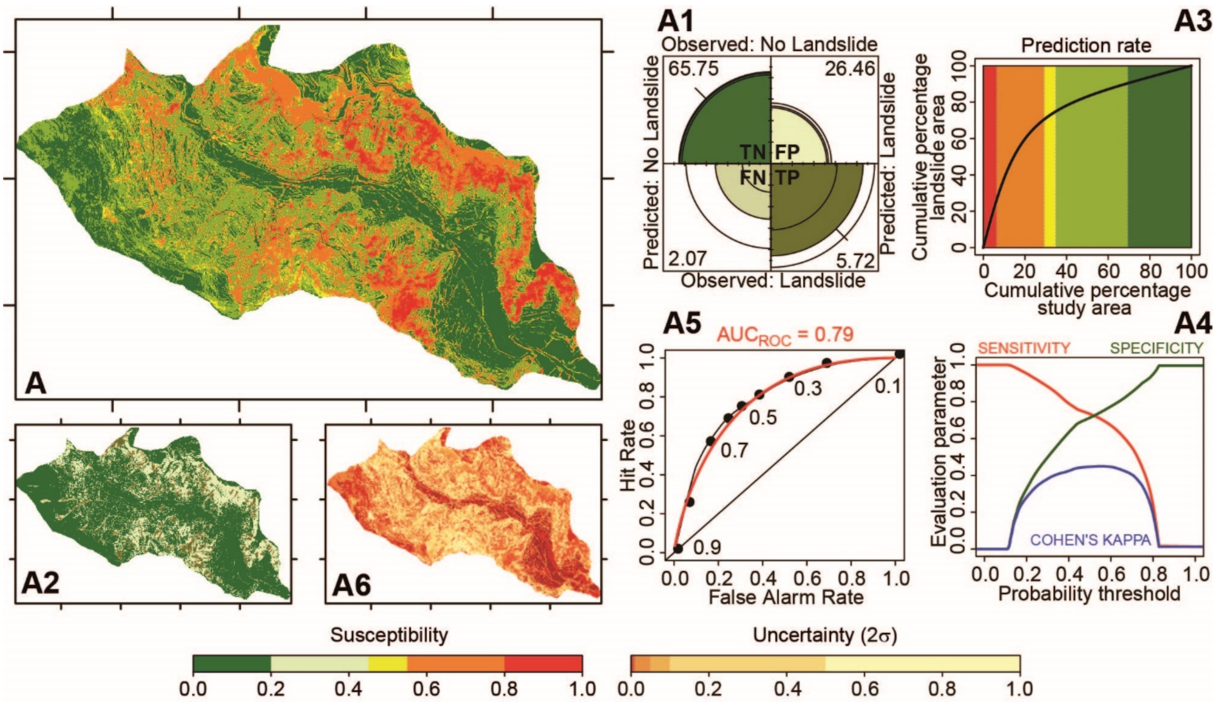


Figure 5. Pixel-based landslide susceptibility map (CM) of the test area (A) classified in five unequally spaced classes (see legend). (A1) fourfold plot summarizing the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN); (A2) map of the distribution of the four categories reported in the fourfold plot; (A3) prediction rate curve; (A4) variation in the model sensitivity, specificity, and Cohen's kappa index; (A5) ROC plot; (A6) map of the geographical distribution of the model error. Maps coordinates and scale bar are shown in Figure 2.

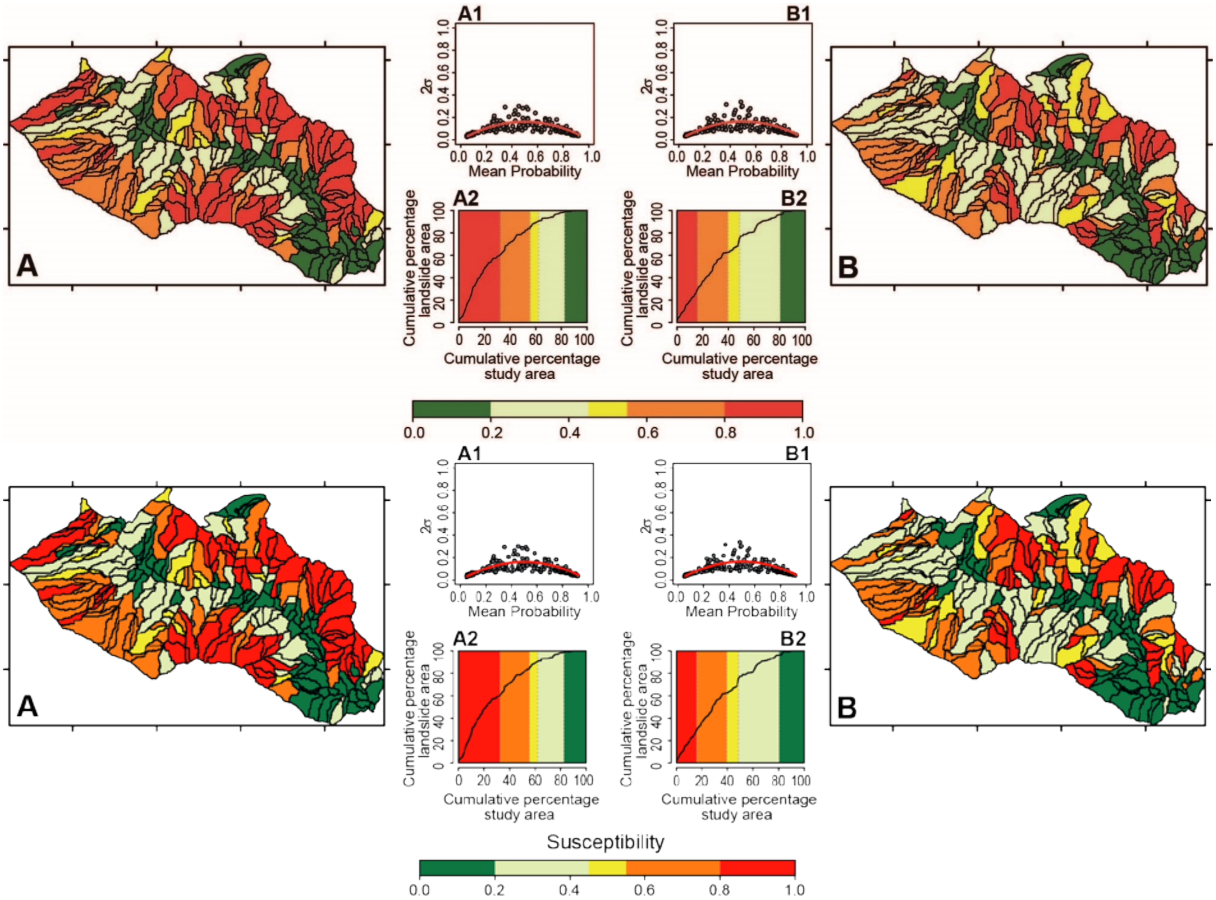
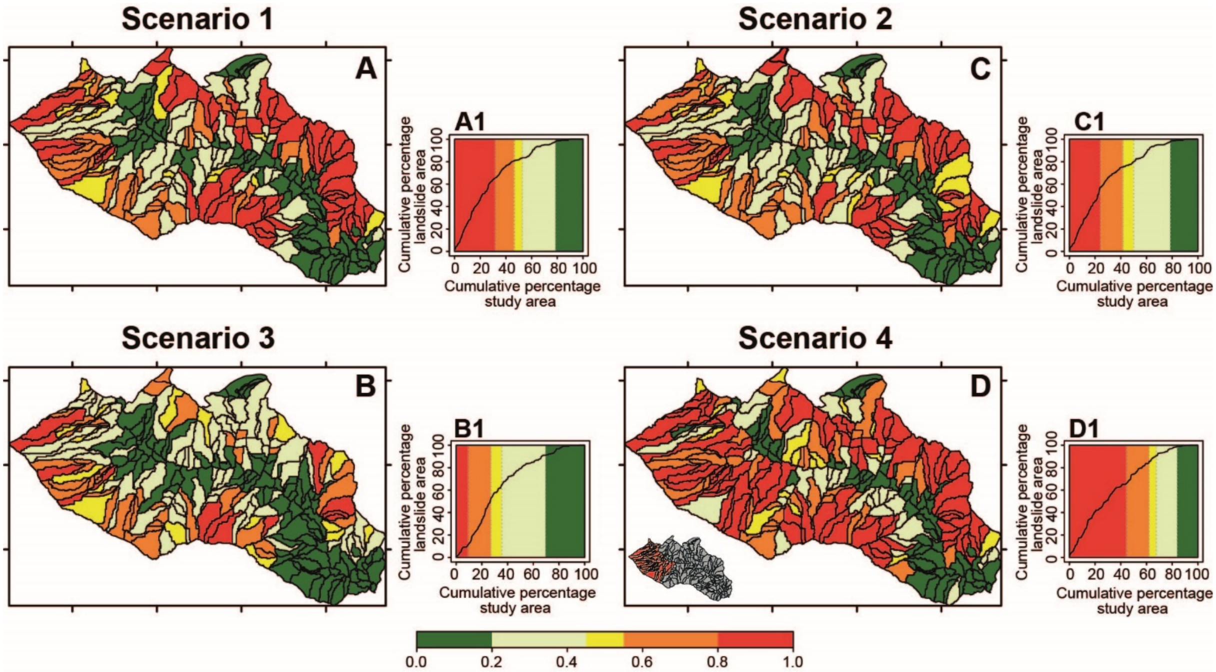


Figure 6. (A) Landslide susceptibility map (CM) prepared using the 2009 land use and (B) using the 1954 land use cover. LS maps are classified in five unequally spaced classes (see legend); (A1, B1) plot showing the model uncertainty estimated in each slope unit; (A2, B2) success rate curves. Maps coordinates and scale bar are shown in Figure 2.

605



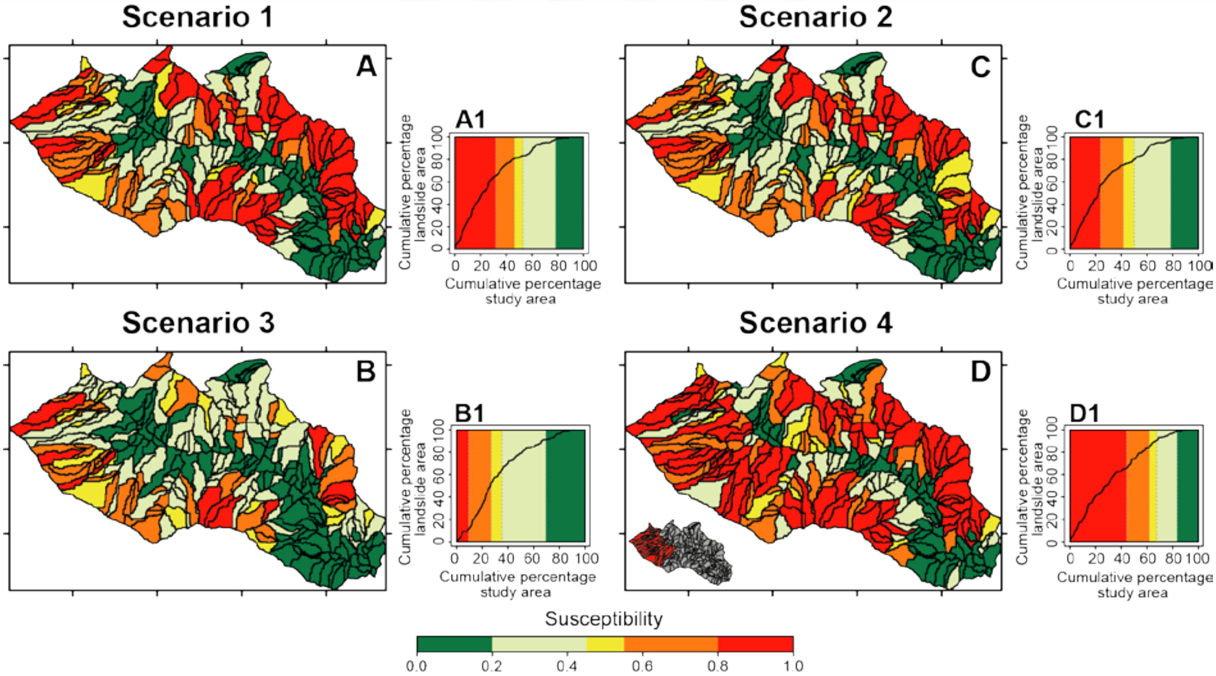


Figure 7. (A, B, C, D) Landslide susceptibility maps (CM) classified in five unequally spaced classes prepared using different land use scenario; (A1, B1, C1, D1) success rate curves. Maps coordinates and scale bar are shown in Figure 2.