Nat. Hazards Earth Syst. Sci. Discuss., 1, 583–623, 2013
www.nat-hazards-earth-syst-sci-discuss.net/1/583/2013/
doi:10.5194/nhessd-1-583-2013

Natural Hazards and Earth System Sciences
Open Access
Discussions

Exploring model sensitivity issues across different scales in landslide susceptibility

F. Catani, D. Lagomarsino, S. Segoni, and V. Tofani

Department of Earth Sciences, University of Florence, Firenze, Italy

Received: 31 January 2013 – Accepted: 9 March 2013 – Published: 21 March 2013

Correspondence to: F. Catani (filippo.catani@unifi.it) and D. Lagomarsino (daniela.lagomarsino@unifi.it)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Despite the large number of recent advances and developments in landslide susceptibility mapping there is still a lack of studies focusing on specific aspects of LSM model sensitivity. For example, the influence of factors of paramount importance such as the survey scale of the landslide conditioning variables (LCVs), the resolution of the mapping unit (MUR) and the optimal number and ranking of LCVs have never been investigated analytically, especially on large datasets.

In this paper we attempt this experimentation concentrating on the impact of model tuning choice on the final result, rather than on the comparison of methodologies. To this end, we adopt a simple implementation of the random forest (RF) classification family to produce an ensemble of landslide susceptibility maps for a set of different model settings, input data types and scales. Random forest is a combination of tree (usually binary) bayesian predictors that permits to relate a set of contributing factors with the actual landslides occurrence. Being it a nonparametric model, it is possible to incorporate a range of numeric or categorical data layers and there is no need to select unimodal training data. Many classical and widely acknowledged landslide predisposing factors have been taken into account as mainly related to: the lithology, the land use, the land surface geometry (derived from of DTM), the structural and anthropogenic constrains. In addition, for each factor we also included in the parameter set the standard deviation (for numerical variables) or the variety (for categorical ones).

The use of random forest enables to estimate the relative importance of the single input parameters and to select the optimal configuration of the regression model. The model was initially applied using the complete set of input parameters then, with progressively smaller subsamples of the parameter space. Considering the best set of parameters we also studies the impact of scale and accuracy of input variables and the of RF model random component on the susceptibility results. We apply the model statistics to a test area in central Italy, the basin of the Arno river (ca. 9000 km^2), we present the obtained results and discuss them.

Results confirm that the choice of parameter set, mapping unit resolution and training sampling method highly influences the overall accuracy of classification and prediction results. This, in turn, implies that a careful sensitivity analysis making use of traditional and new tools should always be performed before producing final susceptibility maps at all levels and scales.

1 Introduction

Landslide susceptibility maps (LSM) are useful documents for land planning, natural risks management and development of mitigation measures. They represent, usually as digital maps, the distributed relative probability of occurrence of a given type of landslide in a given area, without taking into consideration the probability of occurrence in time (Brabb, 1984).

LSMs can be obtained in a variety of manners and a very ample literature is available on the argument, relating on at least 20 yr of history of susceptibility assessment for mass movements. The first method ever adopted is probably the so called heuristic mapping, carried out by a team of expert geomorphologists through the definition of a set of conditioning factors leading to landslide development in a given area on the basis of field surveys and aerial photograph interpretation supported by ancillary map data such as geological maps. This approach, even though subjective, has the advantage of providing a way to exploit the expert knowledge of the geomorphologist and his judgment and has been used in recent times as well (see e.g. Cardinali et al., 2002; Casagli et al., 2004). Unfortunately, though, it is also unavoidably subjective: in an interesting study, for example, Ardizzone et al. (2002) showed that the same area, independently surveyed by 3 teams of geomorphologists, produced 3 very different LSMs, with spatial positioning inconsistencies in the range of 55–65 %.

For this reason many authors started to propose quantitative assessment methods, based on a set of uniquely defined conditioning factors to increase LSM reproducibility and on a variety of weighting techniques to improve accuracy and robustness. Large part of the quantitative methods to produce LSMs relies on regression or classification approaches. The techniques most widely used are probably discriminant analysis (Carrara, 1983; Chung and Fabbri, 1995; Baeza and Corominas, 1996) and logistic regression (Hosmer and Lemeshow 2000; Lee, 2005; Manzo et al., 2012), although other techniques have proved themselves reliable and in some cases more flexible, such as, e.g. Artificial Neural Networks (ANN) (Bianchi and Catani, 2002; Lee et al., 2003, 2004; Ermini et al., 2005; Yilmaz, 2009a), linear regression (Atkinson and Massari, 1998), fuzzy membership (Kanungo et al., 2006), conditional probability or Bayesian methods (Yilmaz, 2010a).

Compared to each other, such methods often seem quite equivalent (Guzzetti et al., 1999; Kanungo et al., 2006; Carrara et al., 2008; Rossi et al., 2010; Yilmaz, 2009b, 2010a) and produce similar results starting from the same input data, even though much depends on the ability of the practitioner in calibrating the various parameters and fine-tuning the model so as to obtain a high-quality result. Recently, on this account, Rossi et al. (2010) suggest that optimal susceptibility predictions might be obtained through the combination of suitable basic LSMs generated by different methods rather than by the application of a single prediction. Several efforts have also been addressed on how to best measure the quality of the LSMs produced by different methods, as well as on what is the influence of mapping errors or mapping choices on the final results. In particular, Frattini et al. (2010) propose a complete framework for the quantitative assessment of LSM quality and also discuss the possible impact of using different methods in terms of cost/benefit analysis. They conclude that ROC (receiver operating characteristic) curves are at present the best quantitative tool to measure LSM quality. As far as mapping errors or model assumptions are concerned, only a few studies are available in the literature trying to get a deeper insight on the actual impact of modelling choices on the final result. Among them, an important contribution has been provided by a study of Guzzetti et al. (2006) that explore the influence of using different types of mapping units for the production of a LSM in Central Italy. They test a discriminant analysis method against an ensemble of 350 different sets of map units, concluding that every LSM product should include such sensitivity analysis in order to obtain a map of the spatial distribution of the estimation error, necessary to complement LSM information. They, in particular, highlights the importance of exploring LSM model calibration and validation. In general, there seems to be a variability of results within an ensemble of single-model runs as high as among different model type runs.

Despite all such efforts, therefore, there is still a lack of studies focusing on specific aspects of LSM model sensitivity. For example, the influence of factors of paramount importance such as the survey scale of the landslide conditioning variables (LCVs), the resolution of the mapping unit (MUR) and the optimal number and ranking of LCVs have never been investigated analytically, especially on large datasets. We have reconstructed and summarised the main lines of LSM model sensitivity in a mental map which is included in the Supplement for reasons of space. Only some of the aspects highlighted in it have been treated in published papers so far.

In this paper, thus, we attempt this experimentation concentrating on the impact of model tuning choice on the final result, rather than on the comparison of methodologies. To this end, we adopt a simple implementation of the random forest (RF) classification family to produce an ensemble of landslide susceptibility maps for a set of different model settings, input data types and scales. RF classification and regression methods offer a very flexible environment for testing model parameters and mapping hypotheses, allowing for a direct quantification of variable importance. The model choice is, in itself, quite innovative since it is the first time that such technique, widely used in remote sensing for image classification (Ham et al., 2005; Pal, 2005), is used in this form for the production of a LSM.

We apply the model statistics to a test area in central Italy, the hydrographic basin of the Arno river (ca. 9000 km^2), we present the obtained results and discuss them. We also use the outcomes of the parameter sensitivity analysis to investigate the different role of environmental factors in the test area.

NHESSD
1, 583–623, 2013

2 Materials and methods

2.1 Random forest classification

As a basic model for LSM we used a random forest implementation based on the tree-bagger object (RFtb) and methods in Matlab. Random forest classification is basically a machine-learning algorithm for non-parametric multivariate classification first developed by Breiman (2001). RF approach are usually adopted in sociological studies (Strobl et al., 2009) and remote sensing image classification (Ham et al., 2005; Pal, 2005). They are being increasingly used, however, also in environmental modelling (Prasad et al., 2006; Strobl et al., 2008; Bachmair and Weiler, 2012). The algorithm exploits random binary trees which use a subset of the parameter space through bootstrapping techniques. Each tree is developed so as to minimize classification errors but the random component influences the results making a single-tree classifier very sensitive to chance in input data selection. For this reason, in place of a single classification or regression tree, the RF-type methods make use of an ensemble of trees (the so-called "forest") thereby ensuring model stability. The RF technique has several advantages with respect to other, more used, multivariate regression or classification methods. Firstly, it does not require assumptions on the distribution of explanatory variables, secondly, it allows for the mixed use of categorical and numerical variables and, thirdly, it is capable of accounting for interrelationship and non-linearities between variables. These are big advantages when working with terrain variables with a high degree of approximation and an intrinsic uncertainty in the assignment to the correct class even in surveyed areas (see e.g. the example of the correct definition of the type of soil for a given non-point location).

A further advantage, even more so for our study, is the ability of RFtb models to provide information on the statistical weight of each single variable on the overall result. This is a direct consequence of the bootstrapping technique inherent to the modelling technique, which excludes or includes variables at every run. This capability can be

NHESSD
1, 583–623, 2013

fruitfully exploited to study the relative importance of the different explanatory variables, a quite important but often neglected aspect of LSM.

In the Matlab implementation of RFtb ("treebagger" method and objects) the relative importance of the LCVs is estimated using the average out-of-bag error (OOBE) in prediction due to the exclusion of the given variable during the bootstrapping phase. The model output is a membership probability to one of the 2 possible classes "landslide" and "no-landslide". The overall performance of the model, instead, can be assessed through the misclassification probability (MP) given by the average classification errors (commission and omission) after a given number of run on a specific RF configuration (see following sections for details).

2.2 Model tests

Before starting the experimentation concerning model parameters scale, accuracy and sampling, we performed a series of tests on the influence of basic RF treebagger (RFtb) settings on the results of modelling.

As we have seen in the previous section, the RFtb implementation requires some preliminary user choices on the forest complexity and on the control over the random component of the model itself.

In particular, we were interested in assessing the stability of the model performances over two important settings: the RFtb tree number and the number of runs required to obtain a consistent prediction of the dependent variable.

2.2.1 Number of trees

A single realization of a RFtb model ends up in a forest structure whose tree number T# is usually established by the user. There is no specific rule nor a best a-priori value for T# but, basically, the more complex the problem the higher the T# value required to build a satisfactory RFtb structure. Simple classification problems require simple structures whilst more complex classification or regression cases are best solved by

NHESSD
1, 583–623, 2013

using higher T# values. In the case of prediction of landslide susceptibility or spatial occurrence we have in general a problem of unknown a-priori complexity. Therefore, in applying a RF-type method, a preliminary exploratory step has to be performed in order to evaluate the optimal T# value ensuring the best cost/benefit ratio.

It is generally known that the higher the T# value the better the overall accuracy of the prediction. It is also known, on the other hand, that high T#-value models require a strong computational effort and larger samples of the studied population.

A technique to find the optimal range of T# is based on the run of a basic model performance test with increasing structure complexity (increasing T#). We applied this technique to the study case plotting the overall accuracy of the prediction (in terms of out-of-bag classification errors OOBE) versus the number of grown trees (the T# value). We propose to choose as the working T# the value at which the OOBE stops decreasing and starts oscillating around a stabilized value acceptable for the prediction (see results section).

2.2.2 Assessment of the random component on results

Given a constant T# value, for each model run, the RFtb method rebuilds a new ensemble, using random choices on the order in which the trees themselves are stored and added to the structure. The RF-type algorithms offer a very robust set of techniques in order to perform this kind of random choice to ensure optimal performances. However, we believe that an assessment of the impact of this randomized choice on the overall model performances is needed to define the best setup for our experiment and to propose a general framework for using RFtb classification and regression methods as a tool for landslide susceptibility estimation in large areas.

For this reason, we performed a series of tests devoted to the evaluation of the noise due to the random nature of the model and of the number of runs that are needed to ensure stability in the model results.

To this end, we compared the model results in terms of out-of-bag errors (OOBE) over different numbers of runs. Using a fixed resolution of 10 m for all the variables

NHESSD
1, 583–623, 2013

(see Sect. 2.5 for a complete description of variable used) we averaged the OOBE for each parameter over 1, 10 and 100 runs and we compared the results in terms of both relative errors and relative variable ranking.

2.3 Parameter set tests

After the first set of tests on the model structure and internal functionality we concentrated on the analysis of the parameter set. In particular, among the many possible issues to be explored, we focused on some of the aspects that are less considered in previous studies for landslide susceptibility: the influence of each single parameter on the final result, the importance of parameter resolution (in terms of pixel size and terrain unit scale), the impact of the dimension of the training set over the final result. In the study, we used a pixel approach over the slope unit concept to allow for a multi-resolution analysis to be carried out. Furthermore, preliminary tests carried out in a parallel research show that the choice of the best method to represent classification units in LSM is dependent on scale but that in general pixel units may be considered a more flexible approach in many cases (Trigila et al., 2013).

2.3.1 Optimal parameter set at different resolutions

An important question, in landslide susceptibility studies, is which and how many conditioning variables (LCV) are needed to optimize the final map.

This is not simply influenced by the total number of variables (LCV#) but also by the resolution at which the analysis is performed using those variables. For this reason we prepared and performed tests where, at different resolutions, a fixed structure RFtb model is repeatedly run with decreasing LCV#. This means running the RFtb estimator with the complete LCV set and then pruning one parameter at a time (the least important in terms of variance explained) in the following runs, so as to reduce the parameter space. We compare the results of each run to find the optimal parameter set in terms of misclassification probability (MP) over contingency tables and parameter OOBE. It

NHESSD
1, 583–623, 2013

is indeed known, in fact, that the larger the parameter space (LCV#), the higher the possibility of describing complex phenomena but at the same time the higher the noise (also in terms of OOBE and AUC overall performances) in the results. This is especially true when we have a limited possibility of increasing the training set along with the LCV#. Larger LCV# would require larger training sets and it is often impossible to have the required number of calibrated and verified samples to carry out a satisfactory model training for large LCV#.

We test this pruning method for 6 different map resolutions to verify which is the influence of scale in parameter ranking and map accuracy in terms of ROC AUC. The mapping unit resolution (MUR) is defined in raster terms as pixel size and we used for this study 10, 20, 50, 100, 250 and 500 m resolutions. It is important to highlight that the original scale of the input data (e.g. DEM or thematic maps) is constant throughout the experiments as we only varied the pixel size of the derived raster maps expressing the spatial distribution of the variable.

2.3.2 Influence of training set dimension on results

Directly connected to the resolution of parameters used in LSM there is the problem of how large the training set should be in order to stabilize statistical predictions. For obvious practical reasons, it is important to find out which is the minimum number of samples (mS#) required to calibrate a model, a quantity which is a function of the dimension of the parameter space LCV#. Usually, the larger LCV#, the larger the mS# needed for model calibration.

The method of sampling the parameter space for building a training set is also important. Mainly, the sampling can be completely random or guided by heuristic rules. In the first case we do not have any control on whose classes or occurrences of a given LCV are sampled whilst in the second case we can constrain the sampling so that every class is represented at least once.

We performed 2 types of tests. In the first, the performance of the model in terms of AUC was analysed using a constant mS# proportion (10 % of study area) with random

NHESSD
1, 583–623, 2013

sampling at different resolutions (MUR = 10, 20, 50, 100, 250 and 500 m). In the second, the same model performance was tested using a constant map resolution (50 m) using variable mS# (from 0.5 % to 50 % of the study area). We also tested the impact of using random versus ordered selection methods for training set sampling.

2.4 Test area and landslide database

The selected test site is the hydrographic basin of the Arno River, Central Italy. The area is 9100 km^2 wide and it is located in the Northern Apennines, a complex thrust-belt system composed by several tectonic units and sedimentary basins. The relief is characterized by a succession of NW–SE ridges (made up of Mesozoic/Tertiary flysches and calcareous units) and Pliocene–Quaternary sedimentary basins.

The Arno basin has a temperate climate with dry summers, November and March are the rainiest months. However, the typical rainfall amounts exhibit strong local differences and the mean annual precipitation ranges from 800 mm on the Chiana valley to 1800 mm on the Apennine ridges.

Landslides are very common in the study area. The geological setting and the lithological characteristics of the area affect the typology and occurrence of landslides, which are mainly constituted by slow-moving deep seated slides (IAEG, 1990; Bertolini et al., 2004; Catani et al., 2005). The majority of the landslides are reactivations of dormant slides and the frequency of first-time landslides is very low; as a consequence the landslide susceptibility chiefly depends on the presence or absence of known instability. To establish the spatial distribution of existing landslides we used a detailed database (Catani et al., 2005), which was recently updated (Rosi et al., 2012), containing more than 27 000 landslides (Fig. 1).

2.5 Input parameters

The choice of the input parameters is a fundamental step in the susceptibility assessment process. While some parameters are extensively used in landslide susceptibility (e.g. lithology and slope gradient), the effectiveness of many others (e.g. higher derivatives of elevation, soil depth, aspect, structural settings,) is still debated and highly depends on the methodology adopted, the physical setting of the study area and the landslide typology (Carrara and Guzzetti, 1995; Baeza and Corominas, 1996; Segoni et al., 2012).

The number of parameters to be adopted is also debated: effective landslide susceptibility assessments have been carried out with just a few parameters (Ohlmacher and Davis, 2003; Dahl et al., 2010; Akgün, 2012; Pereira et al., 2012) as well as with a relevant number of parameters (Guzzetti et al., 1999; Lee et al., 2002; Gorsevski et al., 2006; Lee and Pradhan, 2007; Nefeslioglu et al., 2011; Felicísimo et al., 2012). However, a high number of parameters do not necessarily grant the quality of the results: it can be demonstrated (Pradhan and Lee, 2010; Floris et al., 2011; Manzo et al., 2012) that an increase in the number of model parameters can even worsen the accuracy of the LSM.

Automated procedures of forward selection of variables in landslide susceptibility mapping have been proposed for several techniques (Carrara et al., 2008; Van den Eckhaut et al., 2009; Costanzo et al., 2012). The use of the RF treebagger algorithm can be a valuable tool in assisting the decision on how many (and which) attributes should constitute the optimal configuration of the susceptibility assessment. An initial and expert-driven selection of the input parameters is therefore not necessary in this study as the pruning procedure will automatically sort LCVs according to importance ranking. We initially selected three main kinds of input parameters: morphometric attributes, thematic attributes and rainfall-related attributes.

Morphometric attributes are quantitative parameters used to characterize landforms. All of them can be put in relation with some physical process or can be used as indicators of the presence/absence of landslides. The original resolution of the Arno basin DEM is 10 m. We resampled it at the other resolutions used in this work (20, 50, 100, 250 and 500 m) and, separately for each of them, a series of topographic attributes were extracted with the same pixel size using ArcGIS 9.3. To encompass the spatial

NHESSD
1, 583–623, 2013

variability of the topographic attributes in the modelling, we defined another series of variables: for each morphometric attribute we considered the standard deviation (for numerical attributes) or the variety (for categorical attributes) calculated on a 3 × 3 pixel window.

- Elevation (ELE; ELE_STD). The elevation basically corresponds to the DEM. This parameter is commonly used in landslide susceptibility assessments as different altitudes may be related to different environmental settings (e.g. vegetation, temperature, rainfall regime, etc. . . .) (Dai and Lee, 2002; Costanzo et al., 2012; Felicísimo et al., 2012; Günther et al., 2012; Sabatakakis et al., 2012). The standard deviation of elevation is closely related to the relative relief and can be considered as an indicator of the potential energy for erosion and mass wasting (Oguchi, 1997; Günther et al., 2012; Kayastha et al., 2012).

- Slope (SLO; SLO_STD). Slope angle is one of the most important preparatory factors as it strongly controls the shear forces acting on hillslopes, therefore it has been widely used in LSM (Aleotti and Chowdhury, 1999; Guzzetti et al., 1999). In addition to the value directly derived from the DTM, its standard deviation was used as an indicator of the potential energy for erosion and mass wasting.

- Curvature – Curvature is traditionally used to describe the physical characteristics of an area with respect to erosion and runoff processes (Zeverbergen and Thorne, 1987) and to identify landforms related to landslides bodies (Evans, 1998; Ohlmacher, 2007; Catani et al., 2010). Various kinds of curvature can be computed as the second derivative of the surface topography. In this study we used four kinds of curvature:

  - Curvature s.s. (CUR; CUR_STD): for each pixel, the second derivative of elevation is computed in every direction in a 3 × 3 window.

  - Profile curvature (CPR; CPR_STD) expresses the rate at which the slope gradient changes towards the direction of maximum slope. It affects the acceleration/deceleration of superficial flux and thus the erosion/deposition of hillslope loose material.

  - Plan curvature (CPL; CPL_STD) is calculated orthogonally to the direction of the maximum slope and it can be used to characterize the convergence and divergence of flow and to discriminate between watersheds and hollows streamed by 0 order hydraulic network.
  - Combo curvature (CCU; CCU_VAR). This is a categorical variable obtained by the combination of the values of plan and profile curvature assumed in each pixel. The use of this attribute allows characterizing more accurately the flow across a surface. Profile and planar curvature were reclassified in three classes (concave, flat, convex) using the values −1 and 1 as class breaks. Afterwards, the two rasters were overlaid finding 9 possible curvature combinations, which provide information about the shape of the hillslope.

- Aspect (ASP; ASP_VAR) – Aspect represents the orientation in the space of each pixel composing the landscape. This variable can play a key role in landslide susceptibility as it may influence the exposition of the terrain to different amounts of rainfall and solar radiation, thus conditioning the terrain humidity and the vegetation growth (Guzzetti et al., 1999; Dai and Lee, 2002; Demir et al., 2013). In this study the aspect was used a categorical variable after reclassifying its angular values on the basis of the facing direction with respect to the 8 main cardinal directions.

- Flow Accumulation (FLA; FLA_STD; LFA; LFA_STD) – this attribute expresses the upslope contributing area of each pixel (i.e. the size of the area drained by a specific pixel in the map), which has been used in landslide susceptibility assessments as it can be put in relation with water flux or with potential soil saturation (Catani et al., 2005; Felicísimo et al., 2012; Xu et al., 2013). Because of the wide extension of the study area, FLA values have a very wide range, therefore

NHESSD
1, 583–623, 2013

we introduced in the analysis also the LFA attribute, which was calculated as the logarithm of the flow accumulation.

- TWI (TWI; TWI_STD) – Topographic Wetness Index (TWI) is defined as $\ln(A/\tan\beta)$, where A is the aforementioned upslope contributing area (or flow accumulation) and β is the slope angle (Beven and Kirkby, 1979). This index is commonly used to characterize the spatial distribution of soil moisture, therefore it is commonly used in landslide susceptibility assessments (Devkota et al., 2013; Pereira et al., 2012; Costanzo et al., 2012; Felicísimo et al., 2012)

Thematic attributes were derived by means of GIS analyses from specific thematic maps.

Lithology (LIT; LIT_VAR) – Lithology is largely acknowledged as one of the most important driving variables in susceptibility assessments, since it directly reflects the geomechanical and hydraulic properties of the bedrock and influences the characteristics of the soil coverage (Dai and Lee, 2002; Catani et al., 2005; Costanzo et al., 2012). In this work we used a 1 : 100 000 lithotechnical map of the Arno basin, previously used in other studies (Catani et al., 2005), where all the geological formations are grouped into 8 classes based on their geotechnical properties: cohesive soils; granular soils; indurated rocks; weakly cemented conglomerates and carbonate rocks; weak rocks; marls and compact clays; rocks with pelitic layers; complex (mainly pelitic) units.

Since the alternation of different lithologies may contribute to slope instability, we also included in the modelling an additional variable defined as the variety of the classes in a 3 × 3 pixels window (LIT_VAR variable).

Land Cover (COV; COV_VAR) – landslide susceptibility is highly influenced by the vegetation cover and the use of land may be used to indirectly account for the human interference on hillslopes (Varnes and IAEG 1984; Costanzo et al., 2012; Pereira et al., 2012). A nine classes (urban fabric; crops and permanent cultivations; grasslands; heterogeneous cultivated lands; forests; rangelands; scrublands; wetlands) land use map was devised starting form a 1 : 50 000 scale land cover map (Catani et al., 2005).

NHESSD
1, 583–623, 2013

As for the lithology, the variety of land cover classes in a 3×3 pixel window was included amongst the susceptibility variables as well.

Distance to roads (RDS) – in some circumstances roads can be considered a landslide predisposing factor: heavy traffic may determine vibrations and sudden increase/decrease of stress, while the construction of roads sometimes requires anthropogenic modification to the hillslope profile or to the drainage system such as road-cuts, fills, culverts, ditches, etc. (Collins, 2008; Ramakrishnan et al., 2013). Therefore distance to roads has been successfully used in landslide susceptibility assessments (Devkota et al., 2013; Demir et al., 2013; Feizizadeh and Blaschke, 2013; Ramakrishnan et al., 2013). We included this preparatory factor in our analysis as a continuous variable by calculating the distance of each pixel from an existing 1 : 10 000 scale shapefile of the road network.

Distance to faults (FTS) – faults are widely used as predisposing factors in landslide susceptibility studies (Devkota et al., 2013; Demir et al., 2013) since they can be related to earthquake induced landslides and because they can be associated to a decrease in the strength parameters of the bedrock and to anomalous groundwater conditions. Faults and other relevant tectonic features of the area were extracted from 1 : 100 000 geological maps and a raster was set up in which each pixel assumes the value of the distance to the closest fault.

Distance to rivers (RIV) – the stream network is an important feature in the geomorphological setting of an area and may directly or indirectly be linked to landslides (Devkota et al., 2013; Demir et al., 2013; Feizizadeh and Blaschke, 2013). Similarly to roads and faults, a shapefile of existing streams (Straheler order > 0) was extracted from 1 : 25 000 technical maps and a raster of distances from the hydraulic network was set up.

Rainfall data has been rarely used in landslide susceptibility models (Günter et al., 2012; Sabatakakis et al., 2012; Schicker and Moon, 2012; Feizizadeh and Blaschke, 2013), mainly because rainfall is considered one of the main triggering factors instead of a predisposing factor. It is often assumed that rainfall is related to the temporal

NHESSD
1, 583–623, 2013

occurrence of landslides and not to their spatial distribution (Pereira et al., 2012). However, this assumption can be considered valid over small areas where rainfall characteristics can be considered quite homogeneous, while on large areas different rainfall regimes can be observed. We therefore included in our analysis some attributes related to the rain without inserting actual rainfall measurements but considering the predisposition of the territory to be struck by a rainstorm of a given typology. We defined a series of variables Rp_a_t (RP_100_24; RP_100_72; RP_240_24; RP_300_72; RP_30_1; RP_50_6; RP_600_120) expressing the return period (RP) of a rainstorm characterized by a given total rainfall amount (a, expressed in mm) in a given time lapse (t, expressed in hours). These data was already known for 111 locations corresponding to pluviometric stations distributed across the whole territory and were extended to the whole study area using an Inverse Distance Weighting (IDW) interpolation algorithm

Random value (RND). The risk of introducing so many variables in the modelling is to have a chaotic and instable system; therefore we defined a control variable that assumes for each pixel a random value from 0 to 100. This "random variable" was used as a benchmark to better identify "useless" or "pejorative" variables that could be less effective than a random choice of values.

3 Results

3.1 Model tests results

In order to identify the minimum number of trees required to minimize OOBE, we repeatedly run the training sequence and calculated the OOBE as a function of increasing T#. Fig. 2 shows that the OOBE stabilizes starting from T# = 100. Considering that the calculation time depends on T#, we choose 200 trees as the optimal configuration of the model.

Considering 10 m resolution and the full parameters set, for each LCV the OOBE was calculated, executing 1, 10 and 100 runs. Results show that the value obtained with 1 run is almost always comparable to the other values, falling within the range identified by ±3 standard deviations (with the only expected exceptions represented by RND and LIT_VAR). Furthermore, in order to compare mean values obtained with 10 and 100 runs, a *t*-test was implemented to compare results coming out from different number of runs: the calculated value is always lower than the critical value, so that the null hypothesis that the two values (10 runs and 100 runs averages in Table 1) belong to the same distributions can never be rejected (in Table 1 the corresponding probability value is also shown). We can assert that the number of model runs does not affect the OOBE. In Table 1 we also report the variation coefficient (CV) considering the mean and std value calculated for 10 and 100 runs. The highest values are obtained for RND and LIT_VAR: the first one did not affect the classification result, the last one makes little sense at 10 m resolution being derived from a 1 : 100 000 scale map.

3.2 Parameter set and resolution tests results

In order to find the best LCV parameter set, a fixed structure RFtb model has been run at different resolutions in the test area applying a pruning procedure, i.e. progressively decreasing the LCV#. The optimal configuration of the parameter set expressed in terms of misclassification probability (MP) for each tested resolution is reported in Table 2.

Out of the 35 parameters considered in the analysis, a combination of 24 parameters represents the best configuration for the resolution of 20 m and 50 m, while a combination of 9, 15, 14 and 22 are the best sets for the 10 m, 100 m, 250 m and 500 m, respectively.

For each resolution the parameters were ranked on the basis of their importance. The random variable was invariably discarded during the first step pruning procedure as the least important (and always pejorative) explaining variable. Rainfall return periods are always included in the optimal set at all different scales, whereas some factors, as CCU, TWI and ASP are always excluded; ELE has always a high rank. The random parameter is always discarded, highlighting the good performance of the model. The resolutions at 20 and 50 m have almost identical parameter sets with very similar ranks for each parameter. Considering the 10 m resolution, most DEM-derived parameters, land use and lithology are discarded.

An interesting by-product of the pruning procedure is the ability to quantify the parameter importance for each LCV# set. The Fig. 3 shows how the rank of a single parameter (color scale) changes for each mapping unit resolution (MUR) and for decreasing LCV#: when the parameter is discarded its rank is displayed in gray. It can be noticeable how the rank varies with the number of parameters and depending on the MUR. The white boxes points out the LCV# which resulted in the optimal set for each MUR. Some examples of this plot are presented in the discussion. The complete set of rank-MUR-LCV# plots are included in the Supplement for reasons of space.

As described in the previous section in order to find the best training set in terms of resolution and minimum sample dimension (mS#) we performed two types of tests. The first type, reported in Fig. 4, considered a random sampling of 10 % of the study area at different resolutions. The model performance compared in terms of AUC is highest for 50 m resolution (AUC = 0.88) and is lowest for 10 m resolution (AUC = 0.54).

The second type of test is performed using a variable sample dimension mS# (from 0.5 % to 50 %) at a constant resolution of 50 m. The results displayed in Fig. 5 highlight that, as predictable, the higher is the mS#, the higher is the resulting AUC.

4 Discussion

4.1 Model testing implications

The main results of the model testing phase for what concerns model stability tell us that, basically, RFtb approaches in landslide studies are feasible and robust provided that a RF structure of the suitable complexity is used and a preliminary stability test is performed on a single configuration to check for randomization noise effects. This

NHESSD
1, 583–623, 2013

means that, before searching the optimal parameter set it is necessary to configure the RFtb structure for maximum stability.

The final structure uses a T# = 200 which implies a computational effort easily manageable by the great majority of desktop computers. Applications in different fields often require much larger forest densities (Bachmair and Weiler, 2012). Moreover, the random component of the RFtb algorithm does not seem to compromise in any way model stability over multiple runs, which further simplifies the practical implementation of this approach in LSM.

Using the model configuration derived from the preliminary tests we can safely assume that model performances in the following tests are not influenced by model structure but only depend on the input parameter set (number, typology, scale and accuracy of LCVs).

4.2 Parameter importance and scale issues

For what concerns the tests of RFtb performances in classifying the study area for landslide susceptibility, we can say that, in the main, the optimal set of LCVs to be included in the analysis depend strongly on the scale of analysis, i.e. on the resolution terrain unit of reference (MUR).

The LCV ranking and significance is clearly influenced by terrain unit size or scale (MUR, which is equivalent to resolution throughout the document).

Table 2 shows that optimal parameter configurations change with scale in a notable manner. Several LCVs which are significant at one scale can be completely negligible at another. This is probably connected with the original resolution of the survey, mapping and measurement of each single variable. For example, the slope curvature variables (CPL_STD, CUR_STD, CUR, CPL, CPR) do not seem to have any influence on landslide susceptibility when MUR = 10 m. This could be probably due to the fact that the dimension of slope curvature which is meaningful for landslide occurrence has a scale higher than that. Slope profile or planar convexities/concavities with dimensions limited to 30 m (window dimension used to compute curvature for 10 m resolution) are probably not significant for discriminating landslide versus non-landslide units in the study area because they fail to include both detachment and deposition areas.

The only curvature-related LCV that is never discarded is the standard deviation of profile curvature (CPR_STD), which assumes high ranking in LSM classification for MUR = 500 m (coarser scale) and an average ranking at MUR = 50 and 100 m (medium scales).

This result may have a notable importance in LSM, per se. The most immediate implication might be that RFtb multiscale analysis can reveal important clues concerning the scaling characteristics of the landslide size distribution in a given area. In particular, for the case at hand, the frequency size distribution shows a scaling in agreement with the classical double-Pareto recognized in most of the landslide inventories worldwide (Stark and Hovius, 2001; Guzzetti et al., 2002; Stark and Guzzetti, 2009; Brunetti et al., 2009; Van den Eeckhaut et al., 2007) with exponents 0.4–0.5 (small landslides) and 1.8–1.9 (larger landslides) and a roll-over around $10^4\,m^2$ (Catani et al., 2005; Convertino et al., 2013). That means that only a small percentage of landslides in the Arno basin have an area smaller than the 10 m resolution window size for neighbour DEM analysis ($30 \times 30\,m = 900\,m^2$). Curvature at that scale (and also slope) simply cannot capture the shape of terrain connected to most landslides.

Another interesting issue is related to the lack of importance that the prediction model assigns to the topo-hydrological covariates. The TWI and RIV indexes are always discarded in the optimal configuration at all scales whilst the contributing area (FLA_STD, FLA and their logarithms LFA, LFA_STD) performs only slightly better, being discarded or low-ranked. We believe that this reflects the fact that these variables are more related to earth flows or rapid landslides in low-order channels of the hydrographic network (such as debris-flow) (Montgomery and Dietrich, 1994; Costanzo et al., 2012; Felicisimo et al., 2012; Pereira et al., 2012; Xu et al., 2013) which are almost absent from the historical inventory used.

Conversely, the LCVs connected to the major triggering factor in the area, rainfall, are by far the most important, rivalled only by the elevation derived covariates (ELE,

ELE_STD). The pattern of landslide distribution in the geographic space seems to be strongly connected to the spatial pattern of interpolated return time for rainfall events as shown in Table 2. In particular, the most relevant types of rainfall events which dictate failure distribution are 30 mm in 1 h and 240 mm in 24 h; given the typical rainfall regime of the area, these values can be considered as characteristic of intense rainstorms. The areas which are more subjected to this type of event are also characterized by a higher spatial probability of landslide occurrence. This is quite new in the Arno river basin, where previous studies did not consider triggering factors (Catani et al., 2005). The ranking of rainfall parameters remains stable across scales, which means that the autocorrelation of the Rp variables obtained for the study area on the basis of the rainfall gauge distribution is strong within the range considered in the analysis (max MUR = 500 m). Since the Rp distribution has been obtained starting from sample points with average inter-distance in the order of 10 km, our analysis shows that this is a correct scale to represent meteorological phenomena for what concerns landslide triggering prediction. Present day state-of-the-art weather forecast systems offer a series of predictions at the meso-β-scale (20 to 100 km) or at the meso-γ-scale (2 to 20 km) that for local studies and predictions can be pushed to decametric pixel sizes only making use of statistical downscaling techniques (Mercogliano et al., 2013). This implies that, even though at the moment the rainfall forecast models do not have the resolution needed for accurate real-time deterministic landslide prediction, they have probably a resolution which is suitable for LSM applications.

The low "classification-power" of many other classical LCVs may again be related to scale issues. For example, the LCVs connected to soil type (LIT, LIT_VAR), derived from a geological map at the 1 : 100 000 scale, do not appear to be important at any scale. We believe that the local differences in lithology or soil/rock composition usually represented in the geological maps at that scale are not accurate enough to capture the local factors leading to landslide occurrence or, alternatively, that they do not actually represent the true surface situation in which regolith and soil cover are the real object of mass failure instead of bedrock material. This consideration cannot be underestimated

NHESSD
1, 583–623, 2013

and has important implications on the way geologists should map surface deposits when the final objective is landslide prediction and forecasting.

The strong linkages between LCVs nature and landslide size distribution in a given area are also highlighted in the overall classification results obtained at the different scales. The mapping accuracy, evaluated using the area under the ROC curve (AUC), is again dependent on scale. The Fig. 4 depicts the ROC curves obtained running the RFtb model at a fixed proportion of training area (10 % with a random sampling scheme) carried out at 6 different terrain unit resolutions (10, 20, 50, 100, 250 and 500 m). The best performances are obtained using a MUR = 50 m (AUC = 0.88) and MUR = 100 m (AUC = 0.81). Very poor results are conversely obtained using finer resolutions (MUR = 10 m AUC = 0.54; MUR = 20 m AUC = 0.58). (See maps in Fig. 5 for a visual comparison.) Our hypothesis is that the 50–100 m scale is the one that best represents the compromise between landslide size and LCVs accuracy in the Arno river basin, given the available input data. This implies that, in agreement with Guzzetti et al. (2006) suggestions, our findings underline the need of performing sensitivity analysis whenever an effective LSM has to be produced for land planning or civil protection purposes in a given area.

The Fig. 6 depicts a new type of plot (see also Fig. 3) which can be considered a support tool for sensitivity analysis in LSM. In it, 5 LCVs are considered as an example of what is the effect of MUR on the susceptibility classification power (colour scale for parameter importance) and optimal dataset configuration. We suggest that this plot, or similar graphic tools, might be routinely used to test model performances and to study the sensitivity of susceptibility estimates to model settings before performing statistical predictions on landslide occurrence.

Another important issue emerging from the results is the evident (and expected) influence of the training set dimension (mS#) on the classification performances in terms of AUC (Fig. 7). At the first sight, this would only seem the confirmation of an obvious principle well known to everyone using statistics. However, this aspect of LSM has always been remarkably absent in the majority of published studies, which usually report a unique sample dimension for model training without discussing the implications of changing this constant value. From our results it is clear that no comparison between different LSMs would be possible before carefully assessing the MUR and the mS# used by the modellers. Furthermore, the sampling method itself influences the model performances and classification power (Yilmaz, 2010b). The Fig. 8 illustrates the impact of using two different sampling methods to choose the training set (random versus block selection over the entire study area). For low percentages of sampling (5 % and 10 %) the random choice is by far preferable to the block sampling.

This is even more important because in most of the practical cases of LSM application to the real world of civil protection and risk mitigation, we cannot decide a-priori the mS# and we are forced to use what we have. LSM approaches ensuring high AUC performances only when using mS# higher than e.g. 50 % are useless in areas where only a very low percentage of terrain has been already mapped using an accuracy suitable for model training.

5 Conclusions

We performed a series of tests to understand how model tuning and model parameters can affect landslide susceptibility mapping in a well studied area of Tuscany (the Arno river basin).

We used a specific version of the random forest classification method in which variable importance can be assessed throughout a series of resolutions and parameter sets so as to understand which is the actual impact of model choices on the final result in terms of classification performances.

The main results we have obtained are that the optimal number of parameters varies with scale and resolution and that the importance of each given landslide conditioning variable is influenced by the model settings and the available data. Also, the choice of the training set (both for dimension and location) is of key importance for obtaining accurate results.

NHESSD
1, 583–623, 2013

All the results we have concur to the conclusion that model sensitivity analysis towards tuning choice, parameter sets and scale issues have a paramount importance on LSM and should always be performed before producing maps to be used for effective landslide risk mitigation.

Supplementary material related to this article is available online at: http://www.nat-hazards-earth-syst-sci-discuss.net/1/583/2013/nhessd-1-583-2013-supplement.pdf.

References

Akgün, A.: A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision and likelihood ratio methods: case study at Izmir, Turkey, Landslides, 9, 93–106, 2012.

Aleotti, P. and Chowdhury, R.: Landslide hazard assessment: summary, review and new perspectives, Bull. Eng. Geol. Environ., 58, 21–44, 1999.

Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., and Reichenbach, P.: Impact of mapping errors on the reliability of landslide hazard maps, Nat. Hazards Earth Syst. Sci., 2, 3–14, doi:10.5194/nhess-2-3-2002, 2002.

Atkinson, P. M. and Massari, R.: Generalized linear modeling of susceptibility to landsliding in the central Apennines, Italy, Comput. Geosci., 24, 373–385, 1998.

Bachmair, S. and Weiler, M.: Hillslope characteristics as controls of subsurface flow variability, Hydrol. Earth Syst. Sci., 16, 3699–3715, doi:10.5194/hess-16-3699-2012, 2012.

Baeza, C. and Corominas, J.: Assessment of shallow landslide susceptibility by means of statistical techniques, in: Proceedings of the Seventh International Symposium on Landslides, edited by: Balkema, A. A., Trondheim, Norway, 17–21 June, Vol. 1, 147–152, 1996.

Bertolini, G., Casagli, N., Ermini, L. and Malaguti, C.: Radiocarbon data on Lateglacial and Holocene landslides in the Northern Apennines, Nat. Hazards, 31, 645–662, 2004.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrol. Sci. Bull., 24, 43–69, 1979.

NHESSD
1, 583–623, 2013

Sensitivity issues in landslide susceptibility

F. Catani et al.

Bianchi, F. and Catani, F.: Landscape dynamics risk management in Northern Apennines (Italy), in: Development and Application of Computer Techniques to Environmental Studies, edited by: Brebbia, C. A., and Zannetti, P., 319–328, WIT Press, Southampton, 2002.
Brabb, E. E.: Innovative approaches to landslide hazard mapping, in: Proceedings 4th International Symposium on Landslides, Toronto, 1, 307–324, 16–21 September 1984.
Breiman, L.: Random forests, Mach. Learn., 45, 5–32, 2001.
Brunetti, M. T., Guzzetti, F., and Rossi, M.: Probability distributions of landslide volumes, Nonlin. Processes Geophys., 16, 179–188, doi:10.5194/npg-16-179-2009, 2009.
Cardinali, M., Reichenbach, P., Guzzetti, F., Ardizzone, F., Antonini, G., Galli, M., Cacciano, M., Castellani, M., and Salvati, P.: A geomorphological approach to the estimation of landslide hazards and risks in Umbria, Central Italy, Nat. Hazards Earth Syst. Sci., 2, 57–72, doi:10.5194/nhess-2-57-2002, 2002.
Carrara, A.: Multivariate methods for landslide hazard evaluation, Math. Geol., 15, 403–426, 1983.
Carrara, A. and Guzzetti, F.: Geographical Information Systems in Assessing Natural Hazards, Kluwer Academic Publisher, Dordrecht, The Netherlands, June 1995, 342 pp., 1995.
Carrara, A., Crosta, G. B., and Frattini, P.: Comparing models of debris-flow susceptibility in the alpine environment, Geomorphology, 94, 353–378, 2008.
Casagli, N., Catani, F., Puglisi, C., Delmonaco, G., Ermini, L., and Margottini, C.: An inventory-based approach to landslide susceptibility assessment and its application to Virginio river basin, Italy, Environ. Eng. Geosci., 10, 203–216, 2004.
Catani F., Casagli, N., Ermini, L., Righini, G., and Menduni, G.: Landslide hazard and risk mapping at catchment scale in the Arno River Basin, Landslides, 2, 329–342, 2005.
Catani, F., Segoni, S., and Falorni, G.: An empirical geomorphology-based approach to the spatial prediction of soil thickness at catchment scale, Water Resour. Res., 46, 1–15, 2010.
Chung, C. F. and Fabbri, A. G.: Multivariate regression analysis for landslide hazard zonation, in: Geographical Information Systems in Assessing Natural Hazards, edited by: Carrara, A. and Guzzetti, F., Kluwer Academic, Dordrecht, 107–142, 1995.
Collins, T. K.: Debris flows caused by failure of fill slopes: early detection, warning, and loss prevention, Landslides, 5, 107–120, 2008.
Convertino, M., Morales, F., Troccoli, A., Linkov, I., and Catani, F.: Landslide patterns as fingerprints of climate change: a MaxEnt approach, J. Geophys. Res., in review, 2013.

Costanzo, D., Rotigliano, E., Irigaray, C., Jiménez-Perálvarez, J. D., and Chacón, J.: Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: application to the river Beiro basin (Spain), Nat. Hazards Earth Syst. Sci., 12, 327–340, doi:10.5194/nhess-12-327-2012, 2012.

Dahl, M.-P. J., Mortensen, L. E., Veihe, A., and Jensen, N. H.: A simple qualitative approach for mapping regional landslide susceptibility in the Faroe Islands, Nat. Hazards Earth Syst. Sci., 10, 159–170, doi:10.5194/nhess-10-159-2010, 2010.

Dai, F. C. and Lee, C. F.: Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong Geomorphology, 42, 213–228, 2002.

Demir, G., Aytekin, M., Akgün, A., İkizler, S. B., and Tatar, O.: A comparison of landslide susceptibility mapping of the eastern part of the North Anatolian Fault Zone (Turkey) by likelihood-frequency ratio and analytic hierarchy process methods, Nat. Hazards, 65, 1481–1506, 2013.

Devkota, K. C., Regmi, A. D., Pourghasemi, H. R., Yoshida, K., Pradhan, B., Ryu, I., Dhital, M. R., and Althuwaynee, O.: Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya, Nat. Hazards, 65, 135–165, 2013.

Ermini, L., Catani, F., and Casagli, N.: Artificial neural networks applied to landslide susceptibility Assessment, Geomorphology, 66, 327–343, 2005.

Evans, I. S.: What do terrain statistics really mean?, in: Landform Monitoring, Modelling and Analysis, edited by: Lane, S., Richards, K., and Chandler, J., John Wiley & Sons, Chichester, UK, 119–138, 1998.

Feizizadeh, B. and Blaschke, T.: GIS-multicriteria decision analysis for landslide susceptibility mapping: comparing three methods for the Urmia lake basin, Iran, Nat. Hazards, 65, 2105–2128, 2013.

Felicísimo, A., 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, doi:10.1007/s10346-012-0320-1, 2012.

Floris, M., Iafelice, M., Squarzoni, C., Zorzi, L., De Agostini, A., and Genevois, R.: Using online databases for landslide susceptibility assessment: an example from the Veneto Region (northeastern Italy), Nat. Hazards Earth Syst. Sci., 11, 1915–1925, doi:10.5194/nhess-11-1915-2011, 2011.

NHESSD
1, 583–623, 2013

Sensitivity issues in landslide susceptibility

F. Catani et al.

Frattini, P., Crosta, G., and Carrara, A.: Techniques for evaluating the performance of landslide susceptibility models, Eng. Geol., 111, 62–72, 2010.
Gorsevski, P. V., Gessler, P. E., Foltz, R. B., and Elliot, W. J.: Spatial prediction of landslide hazard using logistic regression and ROC analysis, Trans. GIS, 10, 395–415, 2006.
Günther, A., Reichenbach, P., Malet, J.-P., Van Den Eeckhaut, M., Hervás, J., Dashwood C., and Guzzetti, F.: Tier-based approaches for landslide susceptibility assessment in Europe, Landslides, doi:10.1007/s10346-012-0349-1, 2012.
Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P.: Landslide hazard evaluation: a review of current techniques and their application in a multiscale study, Central Italy, Geomorphology, 31, 181–216, 1999.
Guzzetti, F., Malamud, B. D., Turcotte, D. L., and Reichenbach, P.: Power–law correlations of landslide areas in central Italy, Earth Planet. Sci. Lett., 195, 169–183, 2002.
Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., and Galli M.: Estimating the quality of landslide susceptibility models, Geomorphology, 81, 166–184, 2006.
Ham, J., Chen, Y., Crawford, M., and Ghosh, J.: Investigation of the random forest framework for classification of hyperspectral data, IEEE T. Geosci. Remote Sens., 43, 492–501, 2005.
Hosmer, D. W. and Lemeshow, S.: Applied Logistic Regression, John Wiley and Sons, New York, 2000.
IAEG: Suggested nomenclature for landslides, IAEG Bulletin, 41, 13–16, 1990.
Kanungo, D. P., Arora, M. K., Sarkar, S., and Gupta, R. P.: A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas, Eng. Geol., 85, 347–366, 2006.
Kayastha, P., Dhital, M. R., and De Smedt, F.: Landslide susceptibility mapping using the weight of evidence method in the Tinau watershed, Nepal, Nat. Hazards, 63, 479–498, 2012.
Lee, S.: Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data, Int. J. Remote Sens., 26, 1477–1491, 2005.
Lee, S. and Pradhan, B.: Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models, Landslides, 4, 33–41, 2007.
Lee, S., Choi, J., and Min, K.: Landslide susceptibility analysis and verification using the Bayesian probability model, Environ. Geol., 43, 120–131, 2002.
Lee, S., Ryu, J. H., Lee, M. J., and Won, J. S.: Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea, Environ. Geol., 44, 820–833, 2003.

Lee, S., Choi, J., and Min, K.: Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea, Int. J. Remote Sens., 25, 2037–2052, 2004.

Manzo, G., Tofani, V., Segoni, S., Battistini, A., and Catani, F.: GIS techniques for regional-scale landslide susceptibility assessment: the Sicily (Italy) case study, Int. J. Geogr. Inf. Sci., doi:10.1080/13658816.2012.693614, 2012.

Mercogliano, P., Segoni, S., Rossi, G., Sikorski, B., Tofani, V., Schiano, P., Catani, F., and Casagli, N.: Brief communication: a prototype forecasting chain for rainfall induced shallow landslides, Nat. Hazards Earth Syst. Sci., in preparation, 2013.

Montgomery, D. R. and Dietrich, W. E.: A physically based model for the topographic control on the shallow landsliding, Water Resour. Res., 30, 1153–1171, 1994.

Nefeslioglu, H. A., Gokceoglu, C., Sonmez, H., and Gorum, T.: Medium-scale hazard mapping for shallow landslide initiation: the Buyukkoy catchment area (Cayeli, Rize, Turkey), Landslides, 8, 459–483, 2011.

Oguchi, T.: Drainage density and relative relief in humid steep mountains with frequent slope failure, Earth Surf. Proc. Land., 22, 107–120, 1997.

Ohlmacher, G. C.: Plan curvature and landslide probability in regions dominated by earth flows and earth slides, Eng. Geol., 91, 117–134, 2007.

Ohlmacher, G. C. and Davis, J. C.: Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA, Eng. Geol., 69, 331–343, 2003.

Pal, M.: Random forest classifier for remote sensing classification, Int. J. Remote Sens., 26, 217–222, 2005.

Pereira, S., Zêzere, J. L., and Bateira, C.: Technical Note: Assessing predictive capacity and conditional independence of landslide predisposing factors for shallow landslide susceptibility models, Nat. Hazards Earth Syst. Sci., 12, 979–988, doi:10.5194/nhess-12-979-2012, 2012.

Pradhan, B. and Lee, S.: Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling, Environ. Model. Softw., 25, 747–759, 2010.

Prasad, A. M., Iverson, L. R., and Liaw, A.: Newer classification and regression tree techniques: bagging and random forests for ecological prediction, Ecosystems, 9, 181–199, 2006.

Ramakrishnan, D., Singh, T. N., Verma, A. K., Gulati, A., and Tiwari, K. C.: Soft computing and GIS for landslide susceptibility assessment in Tawaghat area, Kumaon Himalaya, India, Nat. Hazards, 65, 315–330, 2013.

NHESSD

1, 583–623, 2013

Sensitivity issues in landslide susceptibility

F. Catani et al.

Rosi, A., Segoni, S., Catani, F., and Casagli, N.: Statistical and environmental analyses for the definition of a regional rainfall thresholds system for landslide triggering in MURcany (Italy), J. Geogr. Sci., 22, 617–629, 2012.
Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A. C., and Peruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecasts, Geomorphology, 114, 129–142, 2010.
Sabatakakis, N., Koukis, G., Vassiliades, E., and Lainas, S.: Landslide susceptibility zonation in Greece, Nat. Hazards, 65, 523–543, doi:10.1007/s11069-012-0381-4, 2012.
Schicker, R. and Moon, V.: Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale, Geomorphology, 161, 40–57, 2012.
Segoni, S., Rossi, G., and Catani, F.: Improving basin-scale shallow landslides modelling using reliable soil thickness maps, Nat. Hazards, 61, 85–101, 2012.
Stark, C. P. and Guzzetti, F.: Landslide rupture and the probability distribution of mobilized debris volumes, J. Geophys. Res., 114, F00A02, doi:10.1029/2008JF001008, 2009.
Stark, C. and Hovius, N.: The characterization of landslide size distribution, Geophys. Res. Lett., 28, 1091–1094, 2001.
Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., and Zeileis, A.: Conditional variable importance for random forests, BMC Bioinformatics, 9, 307, doi:10.1186/1471-2105-9-307, 2008.
Strobl, C., Malley, J., and Tutz, G.: An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests, Psychol. Methods, 14, 323–348, doi:10.1037/a0016973, 2009.
Trigila, A., Frattini, P., Casagli, N., Catani, F., Crosta, G., Esposito, C., Iadanza, C., Lagomarsino, D., Scarascia Mugnozza, G., Segoni, S., Spizzichino, D., Tofani, V., and Lari, S.: Landslide susceptibility mapping at national scale: the Italian case study, in: Landslide Science in Practice, edited by: Margottini, C., Canuti, P., and Sassa, K., 1, ISBN 978-3-642-31324-0, 2013.
Van Den Eeckhaut, M., Poesen, J., Govers, G., Verstraeten, G., and Demoulin, A.: Characteristics of the size distribution of recent and historical landslides in a populated hilly region, Earth Planet. Sci. Lett., 256, 588–603, doi:10.1016/j.epsl.2007.01.040, 2007.
Van Den Eeckhaut, M., Reichenbach, P., Guzzetti, F., Rossi, M., and Poesen, J.: Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium, Nat. Hazards Earth Syst. Sci., 9, 507–521, doi:10.5194/nhess-9-507-2009, 2009.

Varnes, D. J. and IAEG Commission on Landslides: Landslide hazard zonation – a review of principles and practice, UNESCO, Paris, 63 pp., 1984.
Xu, W., Yu, W., Jing, S., Zhang, G., and Huang, J.: Debris flow susceptibility assessment by GIS and information value model in a large-scale region, Sichuan Province (China), Nat. Hazards, 65, 1379–1392, 2013.
Yilmaz, I.: A case study from Koyulhisar (Sivas – Turkey) for landslide susceptibility mapping by artificial neural networks, Bull. Eng. Geol. Environ., 68, 297–306, 2009a.
Yilmaz, I.: Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat – Turkey), Comput. Geosci., 35, 1125–1138, 2009b.
Yilmaz, I.: Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine, Environ. Earth Sci., 61, 821–836, doi:10.1007/s12665-009-0394-9, 2010a.
Yilmaz, I.: The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks, Environ. Earth Sci., 60, 505–519, 2010b.
Zeverbergen, L. W. and Thorne, C. R.: Quantitative analysis of land surface topography, Earth Surf. Proc. Land., 12, 47–56, 1987.

NHESSD
1, 583–623, 2013

Sensitivity issues in landslide susceptibility

F. Catani et al.

Table 1. Out-of-bag errors (OOBE) relative to the used LCVs for different numbers of model runs. Results show that the number of runs does not seem to influence the LCV impact on classification.

	1 run	10 run			100 run			*t*-test	
	OOBE	meanOOBE	Std dev	CV	meanOOBE	Std dev	CV	*t* value	*p* value
ASP	0.095	0.070	0.058	0.828	0.076	0.078	1.020	−0.319	$< t\ (p = 10\%)$
ASP_VAR	0.471	0.375	0.055	0.146	0.382	0.072	0.189	−0.391	$< t\ (p = 10\%)$
CPL_STD	1.391	1.380	0.072	0.053	1.355	0.084	0.062	1.006	$< t\ (p = 10\%)$
CPR_STD	1.606	1.510	0.081	0.054	1.498	0.114	0.076	0.442	$< t\ (p = 10\%)$
CUR_STD	1.611	1.447	0.119	0.082	1.410	0.104	0.074	0.943	$< t\ (p = 10\%)$
CUR	1.294	1.357	0.115	0.085	1.318	0.118	0.090	1.018	$< t\ (p = 10\%)$
CPL	1.200	1.206	0.058	0.048	1.187	0.099	0.083	1.000	$< t\ (p = 10\%)$
CPR	1.185	1.150	0.113	0.098	1.145	0.101	0.088	0.106	$< t\ (p = 10\%)$
FTS	0.947	1.081	0.096	0.089	1.056	0.095	0.090	0.779	$< t\ (p = 10\%)$
RIV	0.280	0.328	0.052	0.160	0.308	0.070	0.228	1.134	$< t\ (p = 10\%)$
RDS	0.639	0.711	0.066	0.093	0.709	0.067	0.095	0.077	$< t\ (p = 10\%)$
ELE	1.748	1.737	0.112	0.064	1.756	0.099	0.056	−0.511	$< t\ (p = 10\%)$
ELE_STD	1.312	1.378	0.138	0.100	1.360	0.117	0.086	0.398	$< t\ (p = 10\%)$
FLA_STD	1.049	1.088	0.153	0.141	1.134	0.112	0.099	−0.907	$< t\ (p = 10\%)$
FLA	1.090	0.923	0.057	0.062	0.934	0.073	0.079	−0.575	$< t\ (p = 10\%)$
LIT_VAR	−0.005	0.023	0.083	3.578	0.034	0.071	2.126	−0.373	$< t\ (p = 10\%)$
LIT	0.919	0.900	0.044	0.049	0.874	0.052	0.060	1.802	$< t\ (p = 5\%)$
LFA	1.054	0.877	0.059	0.067	0.924	0.080	0.086	−2.380	$< t\ (p = 1\%)$
LFA_STD	0.942	1.000	0.143	0.143	1.034	0.121	0.117	−0.719	$< t\ (p = 10\%)$
RND	−0.044	0.019	0.068	3.641	−0.002	0.074	−35.091	0.921	$< t\ (p = 10\%)$
SLO	1.208	1.201	0.092	0.077	1.219	0.111	0.091	−0.596	$< t\ (p = 10\%)$
SLO_STD	1.354	1.322	0.117	0.089	1.361	0.092	0.068	−0.980	$< t\ (p = 10\%)$
CCU	0.725	0.595	0.048	0.081	0.618	0.079	0.129	−1.405	$< t\ (p = 10\%)$
CCU_VAR	0.172	0.219	0.077	0.353	0.209	0.080	0.385	0.407	$< t\ (p = 10\%)$
TWI	0.665	0.550	0.056	0.102	0.547	0.076	0.139	0.167	$< t\ (p = 10\%)$
TWI_STD	0.693	0.596	0.069	0.116	0.614	0.074	0.121	−0.797	$< t\ (p = 10\%)$
COV	0.806	0.859	0.063	0.073	0.862	0.071	0.083	−0.125	$< t\ (p = 10\%)$
COV_VAR	0.109	0.052	0.058	1.114	0.040	0.067	1.692	0.631	$< t\ (p = 10\%)$
Rp_100_24	1.443	1.435	0.097	0.068	1.471	0.088	0.060	−1.104	$< t\ (p = 10\%)$
Rp _100_72	1.631	1.517	0.071	0.047	1.533	0.097	0.063	−0.704	$< t\ (p = 10\%)$
Rp_24_24	1.500	1.604	0.067	0.044	1.593	0.095	0.060	0.476	$< t\ (p = 10\%)$
Rp_300_72	1.519	1.536	0.078	0.051	1.542	0.091	0.059	−0.221	$< t\ (p = 10\%)$
Rp_30_1	1.559	1.477	0.072	0.049	1.471	0.075	0.051	0.281	$< t\ (p = 10\%)$
Rp_50_6	1.398	1.501	0.082	0.054	1.515	0.088	0.058	−0.501	$< t\ (p = 10\%)$
Rp_600_120	1.425	1.418	0.063	0.044	1.433	0.090	0.062	−0.715	$< t\ (p = 10\%)$

NHESSD
1, 583–623, 2013

Table 2. Optimal configurations of the parameter set for each map unit resolution (MUR = 10, 20, 50, 100, 250 and 500 m). Discarded parameters are marked by –. Numbers represent the rank of each parameter according to OOBE. MP is the overall misclassification probability of the given ensemble. Some of the parameters are never present in the optimal sets.

	MUR = 10 m MP: 0.105	MUR = 20 m MP: 0.110	MUR = 50 m MP: 0.109	MUR = 100 m MP: 0.111	MUR = 250 m MP: 0.104	MUR = 500 m MP: 0.108
ASP	–	–	–	–	–	–
ASP_VAR	–	–	–	–	–	–
CPL_STD	–	15	8	10	–	15
CPR_STD	9	11	7	8	13	2
CUR_STD	–	10	11	12	11	17
CUR	–	8	19	–	–	11
CPL	–	16	16	15	–	16
CPR	–	9	12		14	13
FTS	–	22	20	14	3	14
RDS	–	–	22	–	–	–
RIV	–	–	–	–	–	–
ELE_STD	8	4	6	6	7	18
ELE	2	1	1	4	1	3
FLA_STD	–	19	14	–	–	12
FL A	–	21	24	–	–	21
LIT	–	20	23	–	–	–
LIT_VAR	–	–	–	–	–	–
LFA	–	23	–	–	–	22
LFA_STD	–	18	18	–	–	19
SLO	–	14	15	–	–	20
RND	–	–	–	–	–	–
SLO_STD	–	5	9	9	6	7
CCU	–	–	–	–	–	–
CCU_VAR	–	–	–	–	–	–
TWI	–	–	–	–	–	–
TWI_STD	–	–	–	–	–	–
COV	–	24	21	–	–	–
COV_VAR	–	–	–	–	–	–
Rp_100_24	–	7	13	7	10	8
Rp_100_72	7	12	10	11	12	9
Rp_24_24	1	3	2	1	2	1
Rp_300_72	6	17	17	13	9	10
Rp_30_1h	4	2	3	2	5	4
Rp_50_6	5	6	5	5	8	5
Rp_600_120	3	13	4	3	4	6
Number of parameters	9	24	24	15	14	22

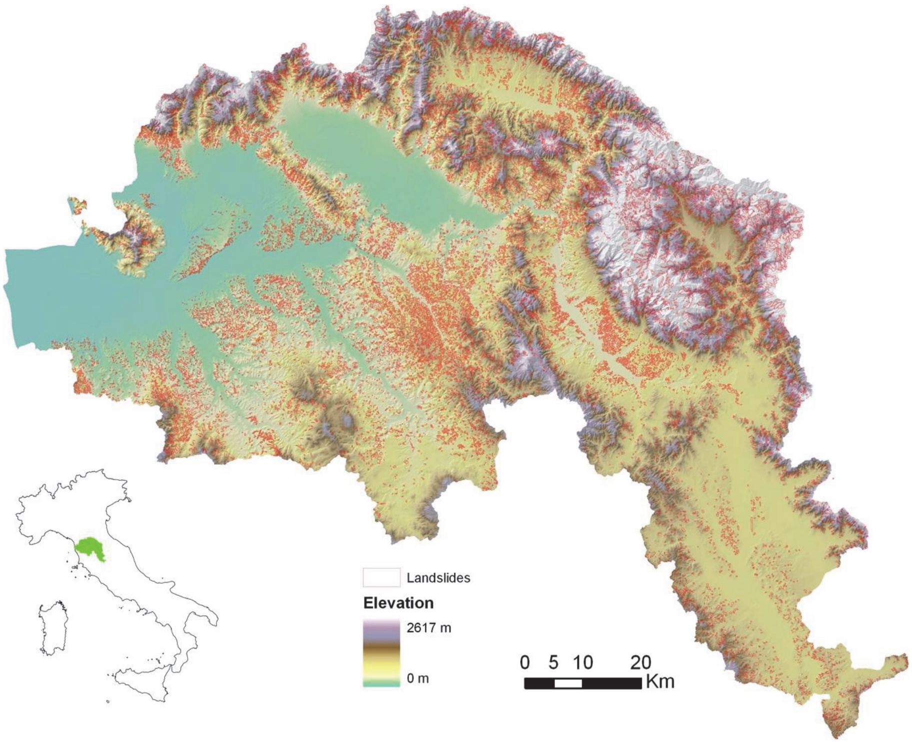

Fig. 1. Location of the study area and distribution of landslides over the Arno river basin (red dots).

NHESSD
1, 583–623, 2013

Sensitivity issues in landslide susceptibility

F. Catani et al.

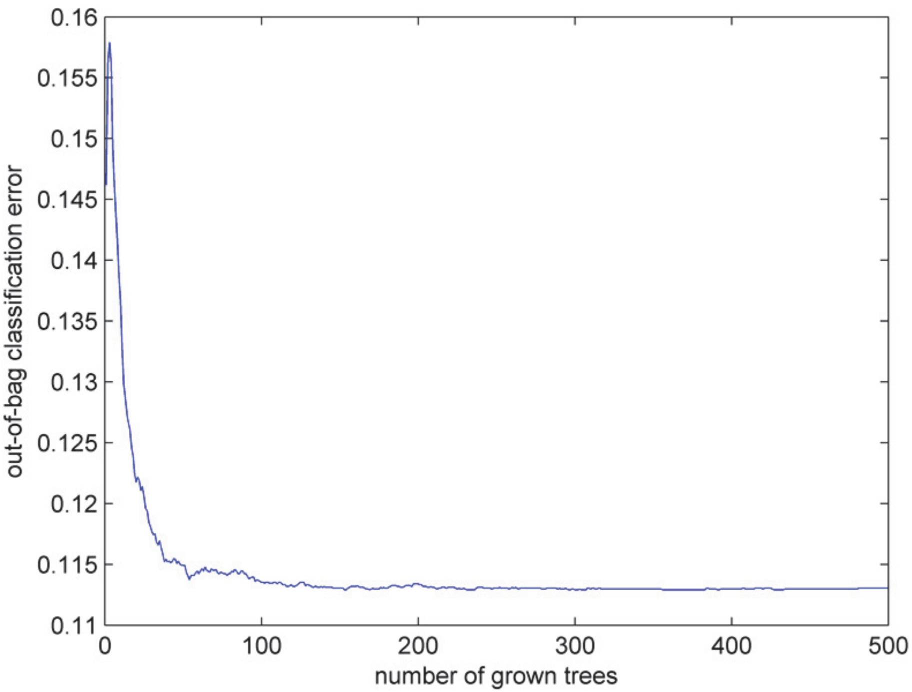

Fig. 2. Plot showing the decrease of OOBE with increasing number of trees T# in the RF structure. A sill value is reached for T# > 100. A working value of T# = 200 was chosen for RFtb model structure used in the tests and experiments.

NHESSD

1, 583–623, 2013

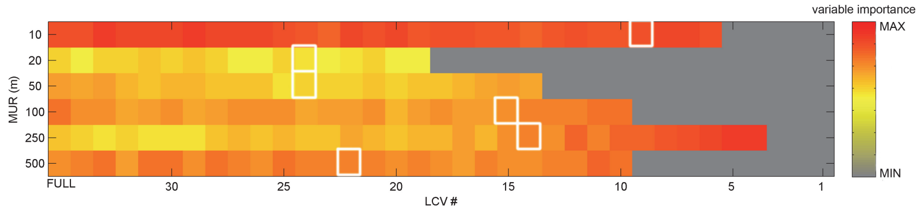

Fig. 3. Rank-MUR-LCV# plot example illustrating the variation of parameter relative importance (expressed as rank using the color ramp on the right) with parameter space (n. of parameters used LCV#) and map unit resolution (MUR in m). Grey colors correspond to combinations of MUR and LCV# in which the parameter importance was estimated as poor or where the parameter was discarded. The white boxes indicate the combination of MUR-LCV# leading to the best classification for each resolution (see Table 2 for optimal set description).

NHESSD
1, 583–623, 2013

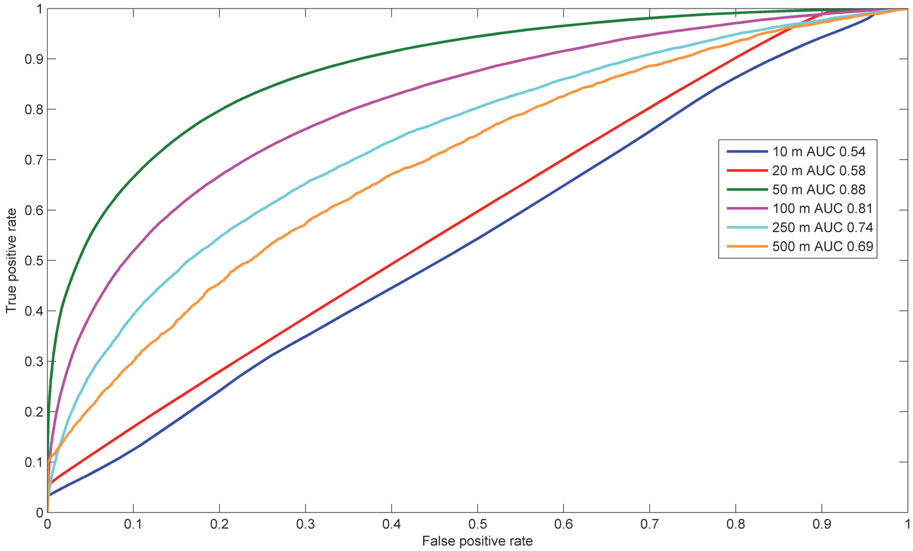

Fig. 4. ROC plots and AUC values for the best classifications obtained at different resolutions. The plots are relative to a model training with mS# = 10 %. The resolution 50 m is the best with AUC = 0.88 whilst no discriminant capability is shown by the RFtb used at 10 and 20 m resolutions. MUR = 100, 250 and 500 m display intermediate accuracies in terms of AUC.

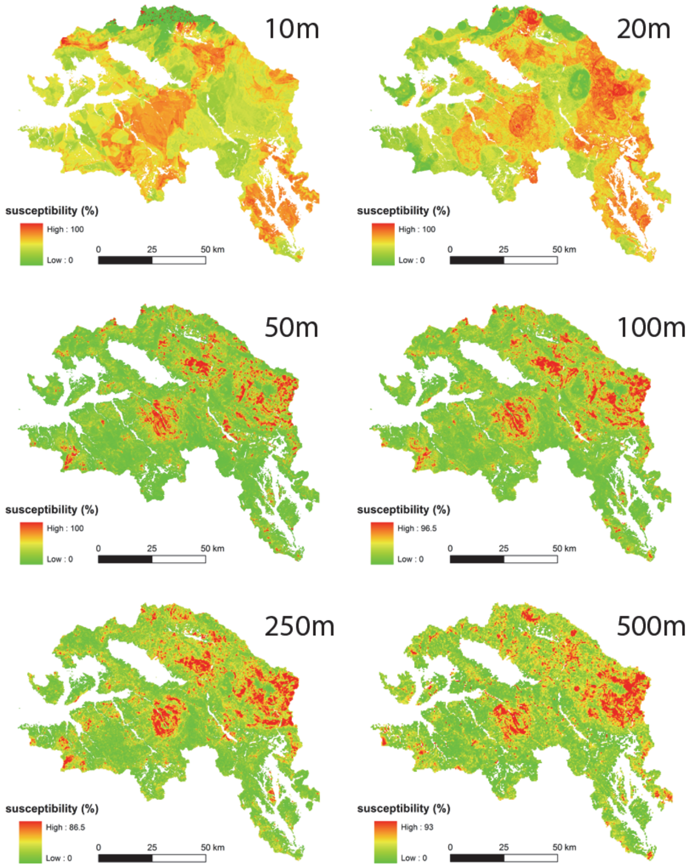

Fig. 5. Susceptibility maps at different resolutions for mS# = 10 % showing the spatial distribution of what is summarized in Fig. 4.

NHESSD
1, 583–623, 2013

Sensitivity issues in landslide susceptibility

F. Catani et al.

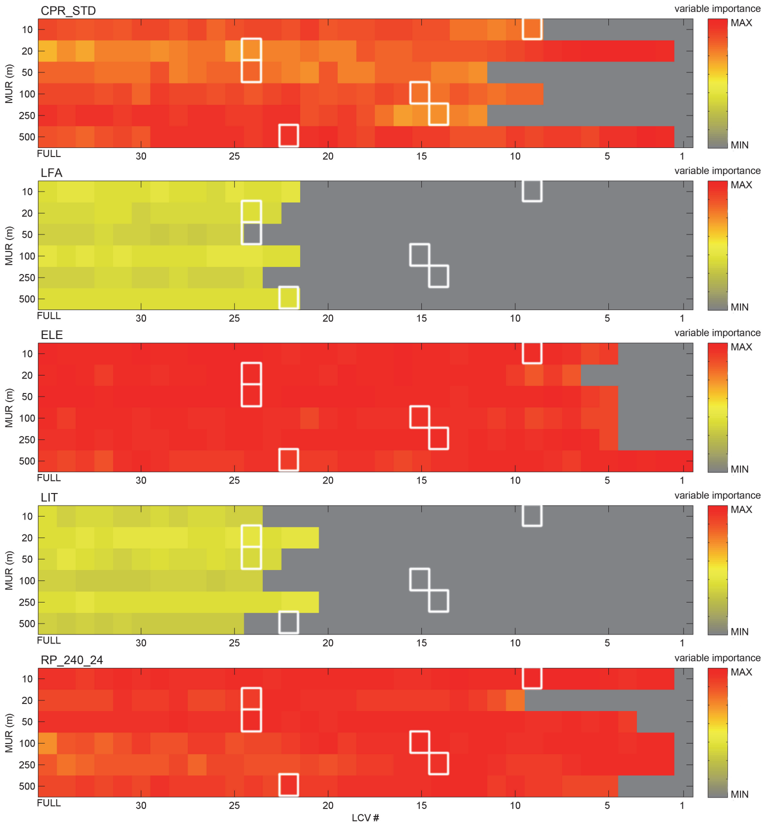

Fig. 6. Rank-MUR-LCV# plots relative to the behavior of 5 LCVs at different resolutions. The plot description is as in Fig. 3. The plots depict visually the efficiency of elevation and rainfall (Rp 240 mm 24 h) in correctly classifying landslide proneness at all scales and almost on every parameter space dimension (LCV#).

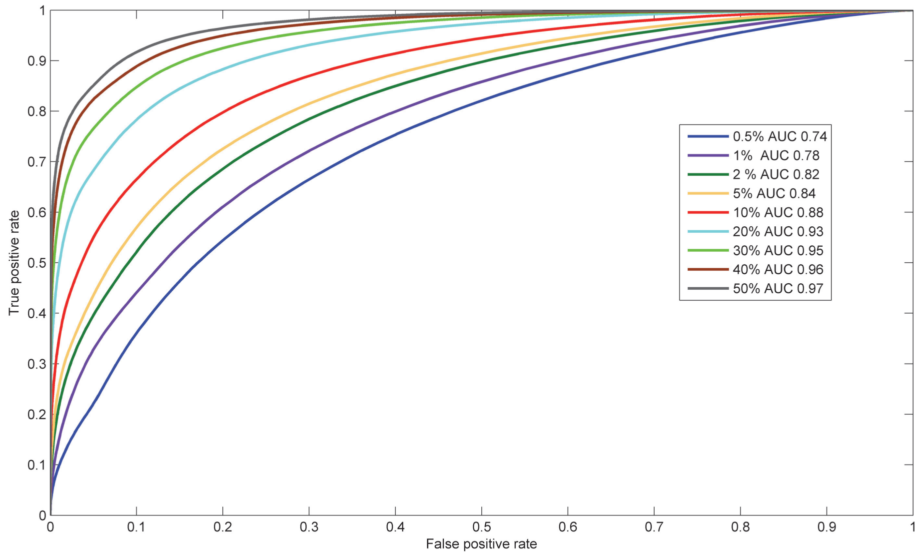

Fig. 7. ROC plots and AUC values for a fixed resolution (MUR = 50 m) at increasing training sample size (0.5 % < mS# < 50 %). Results clearly highlight that AUC increases with mS#, suggesting the need for a research of the best trade-off for such value especially when applying LSM to nearly-unknown areas.

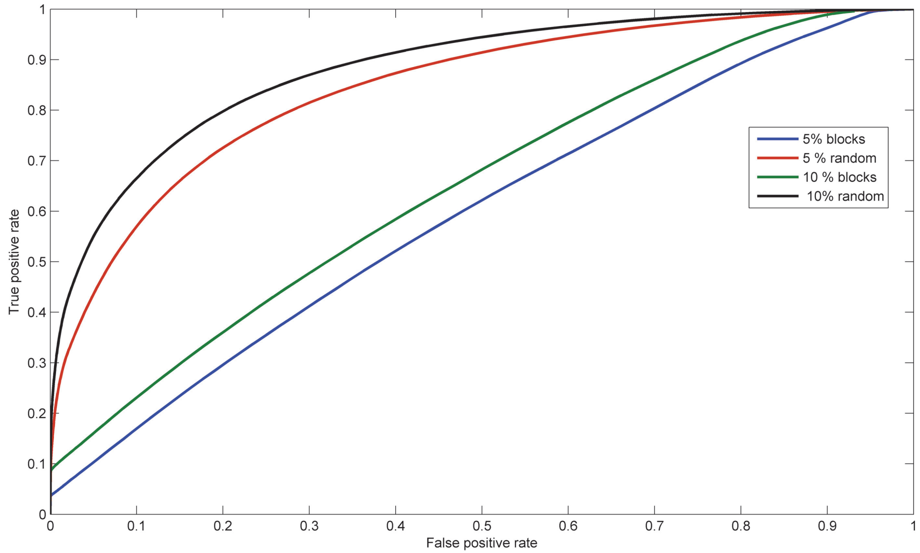

Fig. 8. Influence of the training sampling on the overall classification results. The plot illustrates ROC curves for blocks versus random sampling at 2 different mS# values (5 % and 10 %). The best performances are by far those offered by the random sampling scheme, at least at the tested mS# values.

NHESSD
1, 583–623, 2013