

# Multi-scale Robust Modelling of Landslide Susceptibility: Regional Rapid Assessment and Catchment Robust Fuzzy Ensemble

Claudio Bosco<sup>1</sup>, Daniele de Rigo<sup>2,3</sup>, Tom Dijkstra<sup>4</sup>,  
Graham Sander<sup>1</sup>, and Janusz Wasowski<sup>5</sup>

<sup>1</sup> Loughborough University, Department of Civil and Building Engineering  
Loughborough LE11 3TU, United Kingdom  
[c.bosco@lboro.ac.uk](mailto:c.bosco@lboro.ac.uk)

<sup>2</sup> European Commission, Joint Research Centre,  
Institute for Environment and Sustainability,  
via E. Fermi 2749, I-21027 Ispra (VA), Italy  
[daniele.de-rigo@jrc.ec.europa.eu](mailto:daniele.de-rigo@jrc.ec.europa.eu)

<sup>3</sup> Politecnico di Milano, Dipartimento di Elettronica e Informazione,  
via Ponzio 34/5, I-20133 Milano, Italy

<sup>4</sup> British Geological Survey, Environmental Science Centre Keyworth NG12 5GG,  
United Kingdom

<sup>5</sup> National Research Council (CNR), Bari, Italy

**Abstract.** Landslide susceptibility assessment is a fundamental component of effective landslide prevention. One of the main challenges in landslides forecasting is the assessment of spatial distribution of landslide susceptibility. Despite the many different approaches, landslide susceptibility assessment still remains a challenge. A semi-quantitative method is proposed combining heuristic, deterministic and probabilistic approaches for a robust catchment scale assessment. A fuzzy ensemble model has been exploited for aggregating an array of different susceptibility zonation maps. Each susceptibility zonation has been obtained by applying heterogeneous statistical techniques as logistic regression (LR), relative distance similarity (RDS), artificial neural network (ANN) and two different landslide susceptibility techniques based on the infinite slope stability model. The sequence of data-transformation models has been enhanced following the semantic array programming paradigm. The ensemble has been applied to a study area in Italy. This catchment scale methodology may be exploited for analysing the potential impact of landscape disturbances. At regional scale, a qualitative approach is also proposed as a rapid assessment technique – suitable for application in real-time operations such as wildfire emergency management.

**Keywords:** Landslide susceptibility, Modelling, Ensemble, Semantic Array Programming.

## 1 Introduction

One of the main challenges in modelling landslides is related to the assessment of their spatial probability of occurrence. Estimating landslide spatial probability

may be supported by many different analytical approaches: heuristic, deterministic and statistical [1]. A subset of those approaches may be described as data-transformation models (D-TM) which propagate input data layers and parameters so as to estimate landslide susceptibility as a spatially explicit derived data layer.

Statistical landslide susceptibility methods are based on establishing relationships among measurable variables whose combination is empirically found to correlate with observed landslide occurrences. Theoretical insights guide the selection and analysis of landslide triggering factors (geological, hydrogeological and geomorphological) to determine the most appropriate input – the best suited set of predictors or covariates [2] – to use for statistically reconstructing landslide susceptibility. Domain expertise is therefore essential for supplementing semi-automated computational methods with meaningful semantic heuristics. Estimations are then made of their relative contribution in slope failure and areas are classified as having different hazard or susceptibility degree [3–5]. Multivariate statistical methods are commonly used methodologies for these analyses. This kind of analysis is based upon the presence or not of stability phenomena within the classified areas [6]. The main problem using this approach is the high sensitivity of the results to the input data and the difficulty in deriving the probability of occurrence from the susceptibility [5].

In deterministic approaches, the landslide susceptibility is determined by carrying out a slope stability analysis. It results in computing the distribution of a factor of safety characterizing the study area. The deterministic approaches should be able, in theory, to provide more reliable results (especially where no field measurements are available on landslides and stable areas) but require detailed datasets describing the triggering factors. Many different models, which are usually composed of coupled dynamic hydrological and slope stability models, have been developed by several authors [7–9].

The evaluation of slope stability conditions in a landscape veneer can be performed by considering the local equilibria of forces along pre-determined, shallow slip surfaces representative of translational slide mechanisms. These equilibria can be expressed as a local factor of safety ( $FS$ ), a ratio between the resisting and driving forces acting on a slip surface. If this  $FS$  reaches values lower than unity, the slope will fail.

$$FS = \frac{\text{resisting forces}}{\text{driving forces}} = \frac{\text{shear strength}}{\text{shear stress}} = \frac{S}{\tau}. \quad (1)$$

In this equation  $S$  represents the shear strength mobilised along the slip surface and can be represented, in its simplest form, as:

$$S = c + (\sigma - u) \tan \phi, \quad (2)$$

In this equation  $c$  represents the cohesive resistances in the soil (these resistance may be enhanced by the presence of roots and other elements),  $\sigma$  represents the total stress and  $u$  represents the pore pressure, both acting on the slip surface and affecting the amount of frictional resistance  $\phi$  that can be mobilised.

The driving force is represented by  $\tau$ , the slope parallel component of gravity acting on the soil mass above the slip surface

$$\tau = \gamma Z \cos \beta \sin \beta, \quad (3)$$

where  $\gamma$  is the bulk unit weight and  $Z$  is the depth to the shear surface.

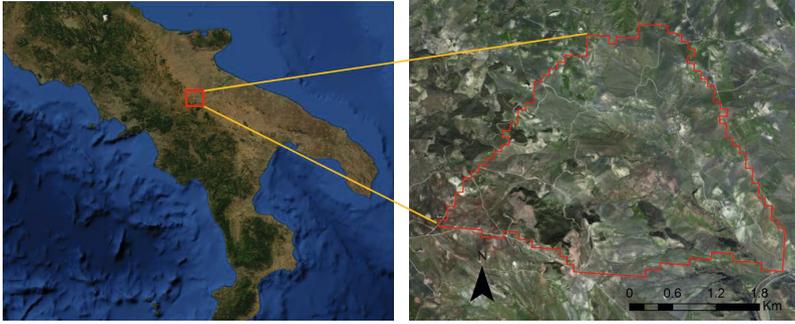
Statistical methods have become popular in recent years [10], including those employed by the British Geological Survey for their GEOSURE product that incorporates an assessment of landslide potential based on an expert elicitation process [11]. One of the main problems in applying these methods is related to changing terrain conditions after a mass movement. Environmental parameters such as slope, land cover or soil depth can be totally different following landslide events. Furthermore, temporal changes in triggering factors in these dynamic landscapes are difficult to evaluate using statistical methods. Deterministic methods may provide more physically based results, but the current limitations due to data quality and availability seriously hampers our ability to simulate landslide processes over relevant spatial and temporal scales [10,12,13]. Some of the soil properties required by deterministic methods, such as soil cohesion or friction angle, are difficult to measure over large areas and show a high spatial variability. Therefore, currently these models still need a high degree of simplification [12]. Because of this limit on present approaches, research is urgently needed for improving landslide susceptibility assessment methodologies.

Landslide susceptibility is also closely linked to many geomorphological and environmental aspects whose perturbation may be driven by natural or anthropogenic changes. This may even affect the landscape/regional scale along with challengingly complex networks of relationships among natural resources and ecosystem dynamics [14–16], also involving climate change and land use alterations [17]. The transdisciplinary analysis of the involved uncertainty and complexity is addressed within the broad domain of integrated natural resources modelling and management (INRMM) [18]. A new approach based on a robust modelling ensemble at catchment scale is here proposed for improving landslide susceptibility estimation. A preliminary regional-scale rapid assessment is also introduced as a possible support in real-time operations such as wildfire emergency management [19,20].

## 2 Methods

### 2.1 Study Area

The study area is located in the Puglia region of southern Italy. The study site (Fig. 1) covers almost 16 km<sup>2</sup> in the municipal territory of Rocchetta Sant'Antonio which is situated in the southern part of the Daunia Appennines. The Daunia Appennines are highly susceptible to landsliding [21,22] and over 20% of the Rocchetta Sant'Antonio area has experienced landslide activity [23].



**Fig. 1.** The study area (Rocchetta Sant'Antonio, Italy). Google Earth, ©2013 Google.

This area has seen a large number of landslides affecting the local economy and has been studied for some years resulting in a database of information (including topography and Digital Elevation Model (DEM), site investigation and geotechnical test data, soil distribution and land use maps) [24–26].

## 2.2 Applied Techniques

Within the study area, precipitation is the main triggering factor for landslide occurrence. In order to improve the spatial prediction of landslides where water is the triggering factor, a combined total of five different deterministic and statistical models have been applied. In order to enhance the determination of landslide susceptibility of the study area, a new method based on an ensemble approach has been used for aggregating the modelling results. The ensemble approach is a reproducible D-TM applied to the results of the array of models and is based on relative-distance similarity (RDS), a flexible technique in INRMM problems [27–29]. The application of an ensemble approach, especially in data-poor regions, could potentially reduce the uncertainty and mitigate local poor performance associated with individual models, by excluding outlier estimations.

Uncertainty may affect these models from the inaccuracy of required inputs data layers  $X$  and parameters  $\theta$  to the approximation of their reconstruction (e.g. by means of other specialised D-TMs) when not directly accessible as available datasets. Uncertainty may also be exacerbated by modelling simplifications or overcomplication. In the latter case, a perhaps theoretically accurate approach might sometimes result in a poorly performing D-TM implementation due to site-specific information gaps and possible impacts of site complexity where multiple conceptual mechanisms coexist as landslide drivers. Furthermore, nontrivial computational models may be affected by *software uncertainty* [18, 30, 31], namely the distance from the theoretical mathematical formulation and the actual model implementation in one or more artificial programming languages.

$Y = f^*(X) = f(\theta^*, X)$	Theoretic D-TM: instance of a model family $f$ by means of selected parameters $\theta^*$
$Y = f^\zeta(\theta^*, \zeta_\theta, X^{\zeta_x})$	Real D-TM with uncertainty affecting input data, parameter calibration and the model itself (software uncertainty [18,30,31])
$Y = \bigcup_{i=1}^n :: \left  f_i^\zeta(\theta^*, \zeta_\theta, X^{\zeta_x}) \right  ::^{sem_i}$	Mitigating uncertainty by ensembling an array of $i \in \{1, \dots, n\}$ heterogeneous models (design diversity [18,30]) with semantic checks $sem_i$ (semantic array programming [31–33])

(4)

In order for the uncertainty to be mitigated, a robust fuzzy ensemble model is proposed to aggregate an array of different susceptibility zonation maps. Each susceptibility zonation has been obtained by applying heterogeneous models, so as to increase design diversity [18,30]. The technique is designed to scale to different arrays of models. Each model is adapted to fit the ensemble array by wrapping its interface so as to behave as a semantically enhanced module. Semantic checks further mitigate inconsistencies between input data, parameters and outputs, following the paradigm of semantic array programming (SemAP) [31–33]. SemAP complements the compactness of array programming notation with an effort towards the most concise generalisation of autonomous tasks as modules which are subject to array-based semantic checks. Each model is considered as a semantically-enhanced module of the ensemble. A few straightforward semantic checks  $sem_i$  are exemplified in the following with the notation  $::sem::$  with hyperlink to the corresponding online description.

Exploiting the available landslide maps and the available environmental information (DEM, land cover and lithological map, geomorphology) two deterministic models based on the infinite slope equation (SINMAP [34, 35] and a simple slope stability model inspired by van Beek’s PROBSTAB model [7]), as well as three statistical models based on Artificial Neural Network (ANN), Logistic Regression (LR) and Relative Distance Similarity (RDS), were calibrated and validated. The theoretical basis of SINMAP involves a mechanistic infinite slope stability model linked with a topographically based steady-state hydrology model. The theory at the basis of the model applies to translational slides where fluctuating pore pressures form the dominant trigger factor. The landslide susceptibility distribution is governed within the model by calculating slope and specific catchment areas starting from a DEM. The model parameters are allowed to be flexible, following a uniform distribution between an upper and lower limit. The parameters may be calibrated using geographic calibration regions based on lithological, land cover or soil characteristics [34, 35].

The implemented slope stability model (hereafter named TransSlide) is based on a translational slope stability function that calculates the factor of safety at

a potential shear plane based on variations of groundwater level and volumetric moisture content. TransSlide is based on eqs. 1, 2 and 3 where:

$$\sigma = Z\gamma \cos^2 \beta \quad (5a)$$

$$u = \gamma_w WL \cos^2 \beta \quad (5b)$$

( $\gamma_w$  is the density of water ( $kNm^{-3}$ ) and WL is the groundwater height above the shear surface). The static inputs include soil shear strength (cohesion and friction) and an additional root cohesion function representing land use. The model functions in a raster-based environment and calculates the local instability on the balance between resisting and driving forces. The paucity of available data in the test region meant that the full functionality of TransSlide could not be mobilised.

Logistic regression analysis, introduced by Cox [36] in 1958, is one of the more commonly used statistical methods in earth science [37]. However, it does not predict presence or absence of landslides, but the probability of occurrence [2]. The predicted values, ranging from 0 to 1, representing the landslides hazard can be expressed as:

$$P(y) = \frac{1}{1+e^{-y}}, \quad (6)$$

where  $P$  is the probability of landslide occurrence. It describes the relationship between a dependent variable ( $y$ ) (assuming a distribution between presence [1] and absence [0]) and  $n$  explanatory variables (the covariates  $C_c^1, C_c^2, \dots, C_c^n$  in the spatial cell  $c$ ) where:

$$y = b_0 + b_1 C_c^1 + b_2 C_c^2 + \dots + b_n C_c^n \quad (7)$$

and  $b_0$  is the intercept of the model with  $b_1, b_2, \dots, b_n$  the regression coefficients.

The maximum-likelihood method is used for estimating the coefficients of the logistic multiple regression model. Because of the non-linearity between independent variables and probability, parameters estimation requires the application of an iterative algorithm [12].

Among the most recent approaches in landslide spatial susceptibility evaluation there are artificial neural network models. An ANN can be considered as a “computational mechanism able to acquire, represent and compute a mapping from one multivariate space of information to another, given a set of data representing that mapping” [38]. A neural network may be able to learn and describe more complex relationships among input covariates than logistic regression could. Actually, a logistic regression model might be seen as a simple special case of neural network. An ANN architecture is characterised by one or more processing elements (nodes or neurons) which propagate weighted inputs received by other nodes [39]. The learning process comes from adjusting the weights between neurons analysing the error between the predicted and target output. The output of a neural network, after the training, is a model that starting from an input dataset is capable of predicting a target value [40].

The RDS is based on the relative distance between two values  $C_1^j$  and  $C_2^j$  of the  $j$ -th nonnegative covariate. The relative distance is a dimensionless number

between 0 (maximum dissimilarity) and 1 (maximum similarity) and is simply the ratio between the minimum and the maximum value:

$$\frac{\min(C_1^j, C_2^j)}{\max(C_1^j, C_2^j)}. \quad (8)$$

The RDS index of a given multi-dimensional point  $c$  with respect to a set  $A$  of reference points (either stable areas or unstable ones) involves the relative distance among the pairs  $\{C_c^j, C_\alpha^j\}$  for each  $\alpha \in A$  and each dimension  $j$  of the  $N^C$  covariates.

## 2.3 Susceptibility Forecasts

### 2.3.1 Data and Explanatory Variables

For running the statistical models a set of calibration parameters was selected from international literature [37, 39, 41]. Seven parameters, commonly assumed as directly or indirectly related to landslide occurrence [37, 39], have been used for calculating the ANN, LR and RDS models. The Topographic Wetness Index (TWI) [42], the slope angle and aspect, the profile and plan curvature have been calculated starting from a digital elevation model with a resolution of 5 meters. The remaining models covariates are the land cover and the DEM itself.

The input parameters for running the deterministic models (root and soil cohesion, internal friction angle, bulk density of the soil, groundwater height, soil depth and the effective recharge rate), have been calculated using different methods. When actual measured data did not exist, estimated values from the literature were applied, if available, otherwise physically sensible values based on our judgment were used.

The recharge rate used in SINMAP (50 mm/d) was derived comparing the available climatological data (Rocchetta SantAntonio and Rocchetta scalo meteorological stations) and approaches from literature [43]. The transmissivity rate required by the model ( $m^2/hr$ ) was calculated using the equation

$$T = K \cdot b, \quad (9)$$

where  $K$  ( $m/hr$ ) is the saturated hydraulic conductivity and  $b$  (m) is the soil depth [43]. The hydraulic conductivity as the soil cohesion and the internal friction angle (used both in SINMAP and TransSlide) were derived from soil texture data using typical soil properties values from literature [44, 45].

The general soil characteristics of the catchment were derived by comparing the lithological map of the study area [23] and field data.

### 2.3.2 Calibration of the Models

The three statistical models were executed using both the same training set and the same set of covariates. The training set consists of 2091 points. The selected areas were subject to **::binary::**<sup>1</sup> classification as stable (0) or not stable (1)

<sup>1</sup> [http://mastrave.org/doc/mtv\\_m/check\\_is#SAP\\_binary](http://mastrave.org/doc/mtv_m/check_is#SAP_binary)

**Table 1.** Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) related to the singles and combined applied models. U and S refer on MAE and RMSE calculated respectively on unstable (U) and stable (S) areas only.

	RDS	ANN	LR	SINMAP	TransSl.	median	ENSEMB.
MAE	0.003	0.44	0.37	0.45	0.51	0.35	0.001
MAE U.	0.002	0.42	0.36	0.61	0.68	0.45	0
MAE S.	0.003	0.45	0.38	0.3	0.34	0.25	0.001
RMSE	0.02	0.47	0.43	0.54	0.58	0.4	0.019
RMSE U.	0.01	0.45	0.4	0.65	0.7	0.47	0
RMSE S.	0.03	0.48	0.46	0.4	0.42	0.32	0.026

on the basis of a dataset presented by [23]. The calibration points represent the 0.6% of the catchment area. These have been selected applying the RDS ensemble technique for analysing the relative distance between the model's covariates within the catchment. Applying this method, for selecting the training and validation points, it is possible to minimize the presence in the training set of areas having similar characteristics.

The multicollinearity of the dataset of **::nonnegative::<sup>2</sup>** **::finite::<sup>3</sup>** predictors was also analysed in order to avoid high correlation between different predictors. This was done for preventing the possibility that small changes in the data can cause an erratic change of the coefficient estimates.

For estimating the best architecture of the ANN the sum of squared errors (SSE) was determined for every combination of weight decay and number of neurons calculated by the ANN model during the network training. The final architecture of the ANN comes from the selection of weight decay and number of nodes minimizing the SSE. The calculation of SSE was then used for evaluating the fitting performance of the models, that is the ability of the model to reproduce the calibration set.

The catchment was split into areas based upon vegetation or lithologic data and calibration of the deterministic models was performed using data from the literature. Due to the paucity of high resolution data for calculating the parameters required by SINMAP and TransSlide only a rough calibration was possible. The low resolution of the available data does not allow the modelling of local variations within the catchment. The applied methodology includes a visual calibration based on expert judgment for adjusting the model's parameters in accordance with landslide observations. The adjustments were necessary for capturing a higher proportion of landslides in areas having a low stability index.

### 2.3.3 Validation of Model Performance

The main objective of this method is to estimate the fitting performance of a model with an independent set of data. Using cross-validation it is possible to estimate how accurately the models perform. An initial set of data was parti-

<sup>2</sup> [http://mastrave.org/doc/mtv\\_m/check\\_is#SAP\\_nonnegative](http://mastrave.org/doc/mtv_m/check_is#SAP_nonnegative)

<sup>3</sup> [http://mastrave.org/doc/mtv\\_m/check\\_is#SAP\\_finite](http://mastrave.org/doc/mtv_m/check_is#SAP_finite)

tioned into two subsets, with the calibration carried out on a training subset (2091 pixels) and the validation performed on the testing subset (682 pixels). For analysing the performance error, mean absolute error (MAE) and root mean square error (RMSE) of the models against validation data were calculated.

Heterogeneous quantities provided as model outputs need to be harmonised in order for the models' performance to be comparable. The **::nonnegative::output** of the two deterministic models has been remapped to the corresponding **::possibility::**<sup>4</sup> values  $\in [0\ 1]$  by means of Piecewise Cubic Hermite Interpolating Polynomials (PCHIP) with the codelet (MATLAB language): `pchip([0 0.5 1 1.25 1.5 10  $\infty$ ], [1 0.8 0.6 0.5 0.4 0 0], output )`.

The results of validation are summarized in table 1. The table shows a better performance of the statistical methods when compared with deterministic approaches. Between the applied models, the ensemble and RDS give the lowest errors. Although the high error rate of some model, the simple models' median is the next best result. Its application as a straightforward unsupervised ensemble might prove useful even where no additional information is available (black box output data).

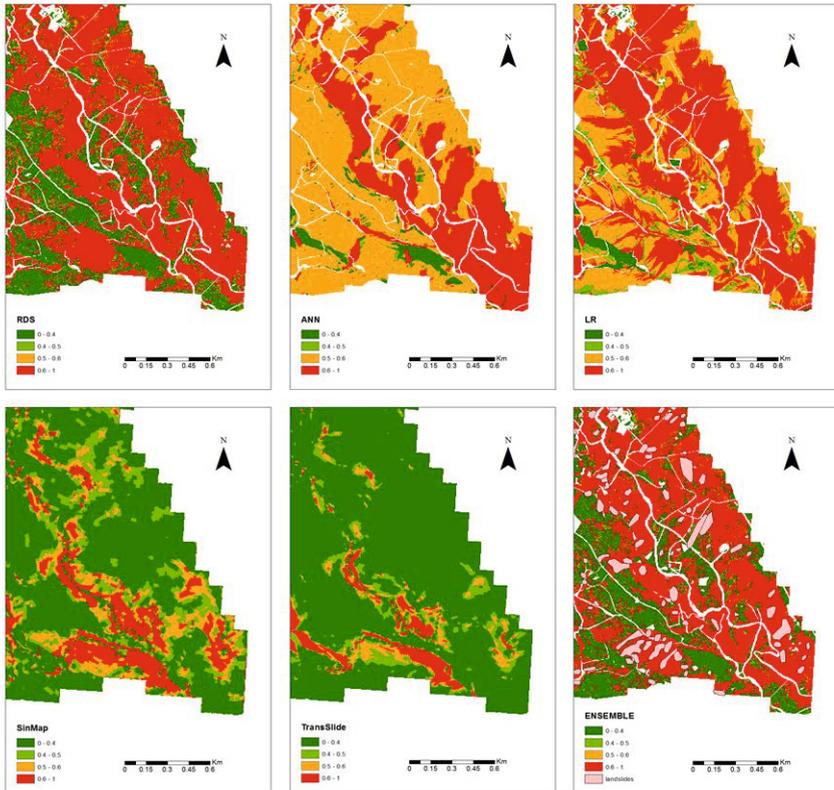
### 3 Results and Discussion

The results of the models are shown in Fig. 2. The classification criteria used for measuring the landslide susceptibility fluctuate in a range between 0 (stable conditions) and 1 (unstable conditions). The adopted classification scheme is as follows: 0-0.4 (stable), 0.4-0.6 (area of model uncertainty) and 0.6-1 (unstable). The better performance of the statistical approaches when compared to deterministic techniques could be linked with both the weakness of data necessary for estimating some of the deterministic model's parameters and the use of spatial data having a sub-optimal resolution.

The proposed ensemble, being a supervised method, slightly improves the best model in the array of outputs (in the worst case, the ensemble would have been equal to the best model in the array). The high performance showed by the RDS approach could be linked with the criterion used for the selection of the training and testing set of data. The possible presence of bias in using a similar technique for selecting the data and calculating the landslide susceptibility need to be further investigated. Because the quality of spatial landslides forecast is largely dependent on the quality of the available datasets, the good performance of the combined model broadens the possibility of applying a quantitative assessment in data poor regions. One of the main difficulties resides in determining how to combine the models for application to a specific area. In landslide susceptibility assessment this can be considered as an unresolved problem [46–48] in [41]. In other disciplines, for example meteorology, a long experience in combining multiple model forecasts already exist [49]. Exploiting ensemble analysis to obtain the most probable forecast is a common practice [41].

---

<sup>4</sup> [http://mastrave.org/doc/mtv\\_m/check\\_is#SAP\\_possibility](http://mastrave.org/doc/mtv_m/check_is#SAP_possibility)

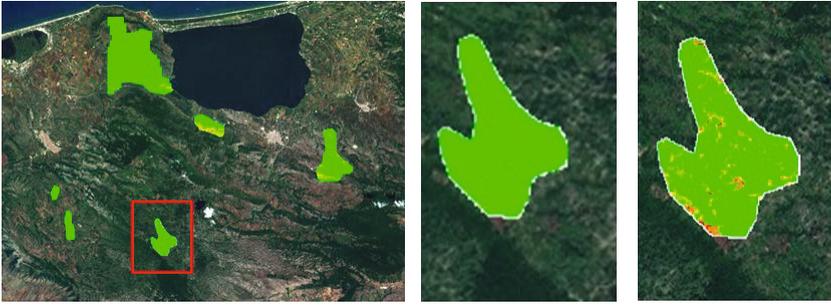


**Fig. 2.** Landslide susceptibility maps of the south-eastern section of the study catchment. The top three diagrams represent the statistical models outputs, the lower left two maps represent the deterministic models. The lower right diagram represents the output of the ensemble modelling.

#### 4 Post-fire Regional Scale Rapid Assessment

European landscapes are affected by many disturbances [50,51]. Among them, wildfires have a major impact especially in southern Europe (Mediterranean region) [52]. Besides their direct environmental impact, wildfires can result in future secondary effects such as shallow landslides and debris flows [53,54].

Applying landslide susceptibility models at a regional scale, in order for the slope stability changes to be rapidly assessed in burnt areas, is a challenging task. Rapid assessment of landslide susceptibility requires fast computational modelling estimations. The availability – for a given spatial extent involving high wildfire risk – of known areas where landslides occurred and of areas of stability cannot be expected. Therefore, empirical models might not be best suited for this kind of analysis. Complex physically based models are also generally inapplicable. However, less data demanding physically based models might provide pre-



**Fig. 3.** Pre and Post-fire shallow landslide susceptibility in Northern Puglia (Italy). Ground layer: Google Earth, ©2013 Google.

liminary assessments of landslide susceptibility. Here, an area of approximately 600 km<sup>2</sup> was analysed calculating the changes in pre and post-fire landslide susceptibility. Six large fires were selected from the European Forest Fires Information System (EFFIS). An increased soil water content and a reduced root cohesion are the main factors inducing shallow sliding after fires [55, 56]. The SINMAP model was applied setting the model parameters for pre and post-fire condition on the basis of literature data [57, 58]. Land cover, aggregate stability and saturated hydraulic conductivity, i.e. the post-fire parameters most susceptible to change significantly [57, 59], were calibrated into the models.

Comparing the pre and post-fire modelling results within the area shown in Fig. 3, indicates that almost 10% of the burnt area changes from stable to unstable conditions. The application of the SINMAP model in post-fire landslide susceptibility analysis has to be considered as a first attempt for applying these techniques at a regional scale. Future research will be carried out to extend the multi-model ensemble architecture from catchment to regional scale, overcoming the possible unavailability of observations on stable and unstable areas.

## 5 Conclusions

Five different techniques for modelling shallow landslide susceptibility were applied to a catchment located in Southern Italy (Rocchetta Sant'Antonio, FG). The stability maps were obtained applying two deterministic models (SINMAP and TransSlide) and three statistical methods (LR, RDS and ANN). A final map was prepared combining the output of the applied models using a robust fuzzy ensemble model. The performance of the models was evaluated against a landslide inventory of the year 2006. The good results of the ensemble model, when compared with the single techniques, make this method suitable to be applied in data poor regions with a lack of calibration and validation data. Because of the uncertainty in selecting a single suitable method for modelling spatial landslide susceptibility in areas characterized by data weakness, the applied ensemble method can potentially result in a less uncertain zonation. Although

these preliminary results are promising, further research is required before this method can be used to communicate the findings with relevant authorities. The landslide susceptibility maps calculated applying the statistical methods were obtained using three different scripts [60] implemented using MATLAB and R languages respectively in GNU Octave and GNU R free software along with the modelling library Mastrave which implements the semantic array programming (SemAP) paradigm. The script applied for calculating the ANN is based on the work of Rossi [41]. Also the scripts used for selecting the training and testing points and for calculating the combined model [60] were written following the SemAP paradigm in MATLAB language, within GNU Octave as computing environment.

**Acknowledgements.** This paper is published with the permission of the Executive Director of the British Geological Survey and the support of the Maieutike Research Initiative.

## References

1. van Westen, C.J., Rengers, N., Terlien, M.T.J., Soeters, R.: Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation. *Geologische Rundschau* 86, 404–414 (1997)
2. Brenning, A.: Spatial prediction models for landslide hazards: review, comparison and evaluation. *Nat. Hazards Earth System Sci.* 5(6), 853–862 (2005)
3. Aleotti, P., Chowdhury, R.: Landslide hazard assessment: summary review and new perspectives. *Bulletin of Eng. Geology and the Env.* 58, 21–44 (1999)
4. Suzen, M.L., Doyuran, V.: A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. *Environmental Geology* 45, 665–679 (2004)
5. Acharya, G.: Analysing the Interactions between Water-induced Soil Erosion and Shallow Landslides. Phd Thesis. University of Canterbury (2011)
6. Van Westen, C.J.: The modelling of landslide hazards using GIS. *Surveys in Geophysics* 21, 241–255 (2000)
7. Van Beek, L.P.H.: Assessment of the Influence of Change in Landslide and Climate on Landslide in Mediterranean Environment. PhD Thesis, Utrecht University, 363p. (2002)
8. Chen, H., Lee, C.F.: A dynamic model for rainfall-induced landslides on natural slopes. *Geomorphology* 51(4), 269–288 (2003)
9. Van Westen, C.J.: Geo-information tools for landslide risk assessment an overview of recent developments. In: *Proceed. of the 9th Int. Symposium on Landslides, Evaluation and Stabilization, Rio de Janeiro*, pp. 39–56 (2004)
10. Van Westen, C.J., Van Asch, T.W.J., Soeters, R.: Landslide hazard and risk zonation; why is it still so difficult. *Bulletin of Engineering geology and the Environment* 65(2), 167–184 (2006)
11. Foster, C., Harrison, M., Reeves, H.J.: Standard and methods of hazard assessment for mass-movements in Great Britain. *Journal for Torrent and Avalanche Control* 166 (2011)
12. Dai, C., Lee, C.F.: Landslide characteristics and slope instability modeling using GIS Lantau Island, Hong Kong. *Geomorphology* 42, 213–238 (2002)

13. Dijkstra, T.A., Dixon, N.: Climate change and slope stability in the UK: challenges and approaches. *Quarterly Journal of Engineering Geology and Hydrogeology* 43(4), 371–385 (2010)
14. Lee, C.F., Ye, H., Yeung, M.R., Shan, X., Chen, G.: AIGIS-based methodology for natural terrain landslide susceptibility mapping in hong kong. *Episodes* 24(3), 150–158 (2001)
15. Karsli, F., Atasoy, M., Yalcin, A., Reis, S., Demir, O., Gokceoglu, C.: Effects of land-use changes on landslides in a landslide-prone area (ardesen, rize, NE turkey). *Environ. Monit. Assess.* 156(1-4), 241–255 (2009)
16. de Rigo, D., Bosco, C.: Architecture of a Pan-European Framework for Integrated Soil Water Erosion Assessment. In: Hřebíček, J., Schimak, G., Denzer, R. (eds.) ISESS 2011. IFIP AICT, vol. 359, pp. 310–318. Springer, Heidelberg (2011)
17. Winter, M.G., Dixon, N., Wasowski, J., Dijkstra, T.A.: Introduction to land-use and climate change impacts on landslides. *Quarterly Journal of Engineering Geology and Hydrogeology* 43(4), 367–370 (2010)
18. de Rigo, D.: Behind the horizon of reproducible integrated environmental modelling at European scale: ethics and practice of scientific knowledge freedom. *F1000 Research*, submitted (exp. 2013)
19. de Rigo, D., Rodriguez-Aseretto, D., Bosco, C., Di Leo, M., San-Miguel-Ayanz, J.: An Architecture for Adaptive Robust Modelling of Wildfire Behaviour under Deep Uncertainty. In: Hřebíček, J., Schimak, G., Kubásek, M., Rizzoli, A. (eds.) ISESS 2013. IFIP AICT, vol. 413, pp. 367–380. Springer, Heidelberg (2013)
20. Di Leo, M., de Rigo, D., Rodriguez-Aseretto, D., Bosco, C., Petroliagkis, T., Camia, A., San-Miguel-Ayanz, J.: Dynamic Data Driven Ensemble for Wildfire Behaviour Assessment: A Case Study. In: Hřebíček, J., Schimak, G., Kubásek, M., Rizzoli, A. (eds.) ISESS 2013, vol. 413, pp. 11–22. Springer, Heidelberg (2013)
21. Iovine, G., Parise, M., Crescenzi, E.: Analisi della franosita nel settore centrale dell Appennino Dauno. *Mem. soc. Geol. Ita.* 51, 633–641 (1996)
22. Magliulo, P., Di Lisio, A., Russo, F., Zelano, A.: Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: A case study in southern Italy. *Natural Hazards* 47(3), 411–435 (2008)
23. Wasowski, J., Lamanna, C., Gigante, G., Casarano, D.: High resolution satellite imagery analysis for inferring surface-subsurface water relationship in unstable slopes. *Remote sensing of Environment* 124, 135–148 (2012)
24. Wasowski, J., Casarano, D., Lamanna, C.: Is the current landslide activity in the Daunia region (Italy) controlled by climate or land use change. In: *Proceedings of the International Conference on Landslides and Climate Change*, Ventor, Isle of Wight, UK, vol. 124, pp. 41–49. Taylor and Francis, Balkema, London, Rotterdam (2007)
25. Mossa, S., Capolongo, D., Pennetta, L., Wasowski, J.: A GIS-based assessment of landsliding in the Daunia Apennines, southern Italy. In: Graniczny, M., Czarnogorska, M., et al. (eds.) *Proceedings of the International Conference Mass Movement Hazard in Various Environments*, Krakow, Poland, vol. 20, pp. 86–91. Polish Geological Institute, Warsaw (2005)
26. Wasowski, J., Lamanna, C., Casarano, D.: Influence of land-use change and precipitation patterns on landslide activity in the Daunia Apennines. *Quarterly Journal of Engineering Geology and Hydrogeology* 43, 387–401 (2010)
27. Bosco, C., de Rigo, D., Dewitte, O., Poesen, J., Panagos, P.: Modelling Soil Erosion at European Scale. Towards Harmonization and Reproducibility (in prep.)

28. de Rigo, D., Guariso, G.: Rewarding Open Science: A Collaborative Review System for Semantically-Enhanced Free Software and Environmental Data Modelling. IFIP AICT (submitted)
29. de Rigo, D., Barredo, J.I., Busetto, L., Caudullo, G., San-Miguel-Ayanz, J.: Continental-Scale Living Forest Biomass and Carbon Stock: a Robust Fuzzy Ensemble of IPCC Tier 1 Maps for Europe. In: Hřebíček, J., Schimak, G., Kubásek, M., Rizzoli, A. (eds.) ISESS 2013. IFIP AICT, vol. 413, pp. 271–284. Springer, Heidelberg (2013)
30. de Rigo, D.: Software Uncertainty in Integrated Environmental Modelling: the role of Semantics and Open Science. *Geophys. Res. Abstr.* 15, 13292+ (2013)
31. de Rigo, D., Corti, P., Caudullo, G., McInerney, D., Di Leo, M., San Miguel-Ayanz, J.: Toward Open Science at the European Scale: Geospatial Semantic Array Programming for Integrated Environmental Modelling. *Geophys. Res. Abstr.* 15, 13245+ (2013)
32. de Rigo, D.: Semantic array programming for environmental modelling: Application of the mastrave library. In: International Environmental Modelling and Software Society (iEMSs) 2012 International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting, pp. 1167–1176 (2012)
33. de Rigo, D.: Semantic Array Programming with Mastrave - Introduction to Semantic Computational Modelling. The Mastrave Project (2012)
34. Pack, R.T., Tarboton, D.G., Goodwin, C.N.: The SINMAP Approach to Terrain Stability Mapping. In: 8th Congress of the International Association of Engineering Geology, Vancouver, British Columbia, Canada (1998)
35. Pack, R.T., Tarboton, D.G., Goodwin, C.N., Prasad, A.: SINMAP 2. A Stability Index Approach to Terrain Stability Hazard Mapping, technical description and users guide for version 2.0. Utah State University (2005)
36. Cox, D.R.: The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society. Series B (Methodological)* 20, 215–242 (1958)
37. 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(6), 1125–1138 (2009)
38. Garrett, J.: Where and why artificial neural networks are applicable in civil engineering. *J. Comput. Civil Eng.* 8, 129–130 (1994)
39. Pradhan, B., Lee, S.: Landslide risk analysis using artificial neural network model focusing on different training sites. *International Journal of Physical Sciences* 4(1), 1–15 (2009)
40. Lee, S., Ryu, J.H., Kim, I.S.: Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea. *Landslide* 4, 327–338 (2007)
41. Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A.C., Peruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecast. *Geomorphology* 114, 129–142 (2010)
42. Beven, K.J., Kirkby, M.J.: A physically based variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin* 24, 43–69 (1979)
43. Witt, A.C.: Using a GIS (Geographic Information System) to Model Slope Instability and Debris Flow Hazards in the French Broad River Watershed, p. 165. Carolina State University, North Carolina (2005)
44. Morgan, R.P.C., Quinton, J.N., Smith, R.E., Govers, G., Poesen, J.W.A., et al.: The European Soil Erosion Model (EUROSEM): Documentation and User Guide. Silsoe College, Cranfield University (1998)

45. Horn, R., Fleige, H.: A method for assessing the impact of load on mechanical stability and on physical properties of soils. *Soil. Till. Res.* 73(1-2), 89–99 (2003)
46. Guzzetti, F., Cardinali, M., Reichenbach, P., Carrara, A.: Comparing landslide maps: a case study in the upper Tiber River Basin, central Italy. *Environmental Management* 25, 247–363 (2000)
47. Guzzetti, F.: *Landslide Hazard and Risk Assessment*. Ph.D. Thesis, Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität, University of Bonn, Germany, p. 389 (2006)
48. Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Galli, M.: Estimating the quality of landslide susceptibility models. *Geomorphology* 81, 166–184 (2006)
49. Collins, M., Knight, S.: Ensembles and probabilities: a new era in the prediction of climate change. *Philosophic. Transact. of the Royal Soc. A*, 1471–2962 (2007)
50. Bengtsson, J., Nilsson, S.G., Franc, A., Menozzi, P.: Biodiversity, disturbances, ecosystem function and management of European forests. *Forest Ecology and Management* 132(1), 39–50 (2000)
51. Seidl, R., Fernandes, P.M., Fonseca, T.F., Gillet, F., Jnsson, A.M., Merganiov, K., et al.: Modelling natural disturbances in forest ecosystems: a review. *Ecological Modelling* 222(4), 903–924 (2011)
52. Rodriguez-Aseretto, D., de Rigo, D., Di Leo, M., Cortes, A., San-Miguel-Ayanz, J.: A data-driven model for large wildfire behaviour prediction in Europe. *Procedia Computer Science* 18, 1861–1870 (2013)
53. Shakesby, R.A., Doerr, S.H.: Wildfire as a hydrological and geomorphological agent. *Earth Sci. Rev.* 74, 269–307 (2006)
54. Cannon, S.H., Gartner, J.E.: Wildfire-related debris flow from a hazards perspective. In: *Debris-flow Hazards and Related Phenomena*, pp. 363–385. Springer, Heidelberg (2005)
55. Swanston, D.N.: Principal soil movement processes influenced by roadbuilding, logging and fire. In: *Proceedings of a Symposium: Forest Land Uses and Stream Environment*, pp. 29–40. Oregon State University, Corvallis (1971)
56. Swanson, F.J.: *Fire and Geomorphic Processes*. In: *Gen. Tech. Rep. USDA For. Serv. WO-26*. pp. 401–420, Washington DC (1981)
57. Neary, D.G., Klopatek, C.C., DeBano, L.F., Ffolliott, P.F.: Fire effects on belowground sustainability: A review and synthesis. *For. Ecol. Manage.* 122, 51–71 (1999)
58. Blesius, L., Weirich, F.: The use of high-resolution satellite imagery for deriving geotechnical parameters applied to landslide susceptibility. In: *ISPRS Hannover Workshop* (2009)
59. Humphreys, F.R., Craig, F.G.: Effects of fire on soil chemical, structural and hydrological properties. In: Gill, A.M., et al. (eds.) *Fire and the Australian Biota*, Canberra. *Aust. Acad. Sci*, ch. 8, pp. 177–200 (1981)
60. de Rigo, D. and Bosco, C.: A multi model robust ensembling architecture for estimating landslide susceptibility. *Free Software and Semantic Array Programming Research* (in prep., 2013)