Evaluating Performances of Simplified Physically Based Models for Landslide Shallow Susceptibility.

Giuseppe Formetta, Giovanna Capparelli and Pasquale Versace

University of Calabria Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica Ponte Pietro Bucci, cubo 41/b, 87036 Rende, Italy (giuseppe.formetta@unical.it, giovanna.capparelli@unical.it, pasquale.versace@unical.it)

Abstract: Rainfall induced shallow landslides can lead to loss of life and significant damage to private and public properties, and transportation systems, etc. Predicting locations that might be susceptible to shallow landslides is a complex task and involves many disciplines: hydrology, geotechnical science, geology, hydrogeology, geomorphology, and statistics. Two main approaches are commonly used: statistical or physically based models. Reliable model applications involve automatic parameter calibration, objective quantification of the quality of susceptibility maps, and model sensitivity analyses. This paper presents a methodology to systemically and objectively calibrate, verify and compare different models and model performance indicators in order to identify and select the models whose behaviors are the most reliable for particular case studies.

The procedure was implemented in a package of models for landslide susceptibility analysis and integrated in the NewAge-JGrass hydrological model. The package includes three simplified physically-based models for landslide susceptibility analysis (M1, M2, and M3) and a component for model verification. It computes eight goodness of fit indices by comparing pixel-by-pixel model results and measurement data. The integration of the package in NewAge-JGrass uses other components such as geographic information system tools to manage input-output processes, and automatic calibration algorithms to estimate model parameters.

The system was applied for a case study in Calabria (Italy) along the Salerno-Reggio Calabria highway, between Cosenza and Altilia. The area is extensively subject to rainfall-induced shallow landslides mainly because of its complex geology and climatology. The analysis was carried out considering all the combinations of the eight optimized indices and the three models. Parameter calibration, verification, and model performance assessment were performed by a comparison with a detailed landslide inventory map for the area. The results showed that the index distance to perfect classification in the receiver operating characteristic plane (D2PC) coupled with model M3 is the best modeling solution for our test case.

Keywords*:* Landslide modelling; Object Modeling System; Models calibration.

1 INTRODUCTION

Landslides are one of the main dangerous geo-hazards worldwide and constitute a serious menace for public safety leading to human and economic losses (Park, 2011). Geo-environmental factors such as geology, land-use, vegetation, climate, and increasing populations may increase the occurrence of landslides (Sidle and Ochiai, 2006). Landslide susceptibility assessments, i.e. the likelihood of a landslide occurring in an area on the basis of local terrain conditions (Brabb, 1984), is not only crucial for an accurate landslide hazard quantification but also a fundamental tool for the environmental preservation and responsible urban planning (Cascini et al., 2005).

Many methods for landslide susceptibility mapping have been developed and can be grouped in two main branches: qualitative and quantitative methods (Glade and Crozier, 2005; Corominas et al., 2014 and references therein).

Qualitative methods, based on field campaigns and expert knowledge and experience, are subjective but necessary to validate quantitative method results. Quantitative methods include statistical and physically based methods. Statistical methods (e.g. Naranjo et al., 1994; Chung et al., 1995; Guzzetti et al., 1999; Catani et al., 2005) use different approaches such as bivariate statistics, multivariate analysis, discriminant analysis, random forest to link instability factors (such as geology, soil, slope, curvature, and aspect) with past and present landslides. Bivariate statistical methods ignore the interdependence of instability factors whereas multivariate analysis is able to statistically consider their interactions. Other data-driven methods for landslide susceptibility analysis include the use of neural networks (Pradhan, 2011; Conforti et al., 2014), support vector machines (Pradhan, 2013 and citations therein), and Bayesian networks (Lee et al., 2002). Deterministic models (e.g. Montgomery and Dietrich, 1994; Lu and Godt, 2008; Borga et al., 2002; Simoni et al., 2008; Capparelli and Versace, 2011; Lu and Godt, 2013) synthesize the interaction between hydrology, geomorphology, and soil mechanics in order to physically understand and predict the location and timing that trigger landslides. These models generally include a hydrological and a slope stability component. The hydrological component simulates infiltration and groundwater flow processes with different degrees of simplification, from steady state (e.g. Montgomery and Dietrich, 1994) to transient analyses (Simoni et al., 2008). The soil-stability component simulates the slope safety factor (FS) defined as the ratio of stabilizing to destabilizing forces. One of the main advantages of data-driven methods for landslide susceptibility is that they can be easily applied in wide areas while deterministic models are in general applied in local analyses. The latter are more computationally expensive and require detailed input data and parameters, which often involve high uncertainty. On the other hand, data-driven methods assume that landslides are caused by the same combination of instability factors overall the study area, whereas deterministic models enable different triggering mechanisms to be understood and investigated.

The results of a landslide susceptibility analysis strongly depend on the model hypothesis, parameter values, and parameter estimation method. Questions regarding the performance evaluation of the landslide susceptibility model, the choice of the best accurate model, and the selection of the best performing method for parameter estimation are still open. Thus, is needed a procedure that facilitates reproducible comparisons between different models and evaluation criteria aimed at the selection of the most accurate models.

Much effort has been devoted to the crucial problem of evaluating landslide susceptibility model performances (e.g. Dietrich et al., 2001; Frattini et al., 2010 and Guzzetti et al., 2006). Accurate discussions about the most common quantitative measures of goodness of fit (GOF) between measured and modeled data are discussed in Bennet et al., (2013), Jolliffe and Stephenson, (2012), Beguería (2006), Brenning (2005) and references therein. We have summarized them in Appendix 1. Usually one of these indices is selected and used as an objective function (OF) in combination with a calibration algorithm in order to obtain the optimal set of model parameters. However, in most cases the selection of the OF is not justified or compared with other options.

The wrong classifications in landslide susceptibility analysis not only risk a loss of life but also have economic consequences. For example locations classified as stable increase their economical value because no construction restrictions will be applied, while the reverse is true for locations classified as unstable.

In this work we propose an objective methodology for environmental model analysis which selects the best performing model based on a quantitative comparison and assessment of model prediction skills. In this paper the methodology is applied to assess the performances of simplified landslide susceptibility models. As the procedure is model independent, it can be used to assess the ability of any type of environmental model to simulate natural phenomena.

Unlike previous applications, our methodology aims to objectively: i) select a set of the most appropriate OFs in order to determine the best model parameters; ii) compare the performance of a model using the parameter sets selected in the previous step in order to identify the OFs that provides particular and not redundant information; iii) perform a model parameter sensitivity analysis in order to understand the relative importance of each parameter and its influence on the model performance. The methodology enables the user to: i) identify the most appropriate OFs for estimating the model parameters and ii) compare different models in order to select the best one that estimates the landslide susceptibility of the study area.

The procedure is implemented in the open source and GIS based hydrological model, denoted as NewAge-JGrass (Formetta et al., 2014) which uses the Object Modeling System (OMS, David et al., 2013) modeling framework. OMS is a Java based modeling framework whch promotes the idea of programming by components. It provides the model developers with many features such as: multithreading, implicit parallelism, models interconnection, and a GIS based system.

The NewAge-JGrass system, Fig. 1, contains models, automatic calibration algorithms for model parameter estimation, and methods for estimating the goodness of the models prediction. The open source GIS uDig (http://udig.refractions.net/) and the uDig-Spatial Toolbox (Abera et al., (2014), https://code.google.com/p/jgrasstools/wiki/JGrassTools4udig) are used as a visualization and input/out data management system. The OMS framework has been previously used as the core for landslides modeling (Formetta et al., 2016; Formetta et al., 2015). These studies deal with real time early warning systems for landslide risks and involve 3D physically based hydrological modeling of very small catchments (up to around 20 km^2). In contrast, the current application focuses on wider areas landslide susceptibility assessments using completely different physically based models which are presented in the next section.

The methodology presented in this paper for landslide susceptibility analysis (LSA) represents one model configuration within the more general NewAge-JGrass system. It includes two new models specifically developed for this paper: mathematical components for landslide susceptibility mapping and procedures for landslides susceptibility model verification and selection. The LSA configuration also uses two models that have already been implemented in NewAge-JGrass: the geomorphological model set-up and the automatic calibration algorithms for model parameter estimation. All the models used in the LSA configuration are presented in Fig. 1, encircled with a dashed red line.

The methodology is presented in section 2. It was setup considering three different landslide susceptibility models, eight GOF metrics, and one automatic calibration algorithm. The flexibility of the system enables more models, and GOF metrics to be added, and different calibration algorithms can be used. Thus deferent LSA configurations can be created depending on: the landslide susceptibility model, the calibration algorithm, and the GOFs selected by the user. Finally, Section 3 presents a case study of landslide susceptibility mapping along the A3 Salerno-Reggio Calabria highway in Calabria, which illustrates the capability of the system.

2 MATERIALS AND METHODS

2.1 Modelling Framework

The landslide susceptibility analysis (LSA) is implemented in the context of NewAge-JGrass (Formetta et al., 2014), an open source large-scale hydrological modeling system. It models the whole hydrological cycle: water balance, energy balance, snow melting, etc. (Figure 1). The system implements hydrological models, automatic calibration algorithms for model parameter optimization, and evaluation, and a GIS for input output visualization, (Formetta et al., 2011, Formetta et al., 2014). NewAge-JGrass is a component-based model, Each hydrological process is described by a model (energy balance, evapotranspiration, run off production in figure 1). Each model implements one or more components (considering for example the model evapotranspiration in Figure 1, the user can select between three different components: Penman-Monteith, Priestly-Taylor, and Fao). In addition each component can be linked to the others and executed at runtime, this building a model configuration. Figure 1 offers a complete picture of the system and the integration of the new LSA configuration encircled with dashed red lines. More precisely the LSA in the current configuration includes two new models: a landslides susceptibility model and a verification and selection model. The first includes three components proposed in Montgomery and Dietrich, 1994, Park et al., 2013, and Rosso et al., 2006, the latter includes the “three step verification procedure” (3SVP), presented in Section 2. The LSA configuration also includes another two models previously implemented in the NewAge-JGrass system: i) the Horton Machine for geomorphological model setup which computes input maps such as slope and total contributing area and which displays the model’s results, and ii) the particle swarm for automatic calibration. Subsection 2.1 presents the landslide susceptibility model and 2.2 presents the model selection procedure (3SVP).

2.2 Landslide susceptibility models

The landslide susceptibility models implemented in NewAge-JGrass and presented in a preliminary application in Formetta et al. (2015) consist of the Montgomery and Dietrich (1994) model (M1), the Park et al. (2013) model (M2) and the Rosso et al. (2006) model (M3). The three models derive from simplifications of the infinite slope equation (Grahm, 1984, Rosso et al., 2006, Formetta et al., 2014) for the factor of safety:

$$FS = \frac{C \cdot (1+e)}{\left[G_s + e \cdot S_r + w \cdot e \cdot \left(1 - S_r\right)\right] \cdot \gamma_w \cdot H \cdot \sin\alpha \cdot \cos\alpha} + \frac{\left[G_s + e \cdot S_r - w \cdot \left(1 + e \cdot S_r\right)\right]}{\left[G_s + e \cdot S_r + w \cdot e \cdot \left(1 - S_r\right)\right]} \cdot \frac{\tan\varphi'}{\tan\alpha} \quad (1)$$

where FS [-] is the factor of safety, $C=C'+C_{root}$ is the sum of C_{root}, the root strength [kN/m^2] and C' the effective soil cohesion [kN/m^2], φ' [-] is the internal soil friction angle, H is the soil depth [m], α [-] is the slope angle, γ_w [kN/m^3] is the specific weight of water, and w=h/H [-] where h [m] is the water table height above the failure surface [m], G_s [-] is the specific gravity of soil, e [-] is the average void ratio and S_r [-] is the average degree of saturation.

The model M1 assumes a hydrological steady-state, flow occurring in the direction parallel to the slope and neglect cohesion, degree of soil saturation and void ratio. It computes w as:

$$w = \frac{h}{H} = \min\left(\frac{Q}{T}\cdot\frac{TCA}{b\cdot\sin\alpha}, 1.0\right) \quad (2)$$

where T [L^2/T] is the soil transmissivity defined as the product of the soil depth and the saturated hydraulic conductivity, b [L] is the length of the contour line. Substituting eq. (2) in (1) the model is solved for Q/T assuming FS=1 and stable and unstable sites are defined using threshold values on log(Q/T) (Montgomery and Dietrich, 1994).

Unlike M1, the model M2 considers: i) the effect of the degree of soil saturation (S_r [-]) and void ratio (e [-]) above the groundwater table and ii) the stabilizing contribution of the soil cohesion. The model output is a map of safety factors (FS) for each pixel of the analyzed area.

The component (M3) considers both the effects of rainfall intensity and duration on the landslide triggering process. The term w depends on rainfall duration and is obtained by coupling the conservation of mass of soil water with the Darcy's law (Rosso et al., 2006) providing:

$$w = \begin{cases} \dfrac{Q}{T}\cdot\dfrac{TCA}{b\cdot\sin\alpha}\cdot\left[1-\exp\left(\dfrac{e+1}{e\cdot(1-S_r)}\cdot\dfrac{t}{T}\cdot\dfrac{TCA}{b\cdot\sin\alpha}\cdot H\right)\right] & if\ \dfrac{t}{T}\cdot\dfrac{TCA}{b\cdot\sin\alpha}\cdot H \le -\dfrac{e\cdot(1-S_r)}{1+e}\cdot\ln\left(1-\dfrac{T\cdot b\cdot\sin\alpha}{TCA\cdot Q}\right) \\ 1 & if\ \dfrac{t}{T}\cdot\dfrac{TCA}{b\cdot\sin\alpha}\cdot H > -\dfrac{e\cdot(1-S_r)}{1+e}\cdot\ln\left(1-\dfrac{T\cdot b\cdot\sin\alpha}{TCA\cdot Q}\right) \end{cases} \quad (3)$$

These models are suitable for shallow translational landslides controlled by groundwater flow convergence. Shallow landslides usually have a very low ratio between the maximum depth (D) and the length (L) of scar (D/L<0.1, Casadei et al., 2003), involve a small volume of the colluvial soil mantle and present a generally translational failure mechanism (Milledge et al., 2014).
Each component has a user interface which specifies the input and output. Model inputs are computed in the GIS uDig integrated in the NewAge-JGrass system by using the Horton Machine package for terrain analysis (Abera et al., 2014). Model output maps are directly imported in the GIS and are available for the user's visualization.
The models that we implemented present an increasing degree of complexity in terms of the theoretical assumptions for modeling landslide susceptibility. Moving from M1 to M2, the soil cohesion and soil properties were considered, and moving from M2 to M3 rainfall of finite duration was used.

2.3 Automatic calibration and model verification procedure

In order to assess the models' performance we developed a model that computes the most common indices for assessing the quality of a landslide susceptibility map.
These indices are based on a pixel-by-pixel comparison between the observed landslide map (OL) and predicted landslides (PL). They are binary maps with positive pixels corresponding to "unstable" ones, and negative pixels that correspond to "stable" ones. Therefore, four types of outcomes are possible for each cell. A pixel is a true-positive (tp) if it is mapped as "unstable" both in OL and in PL, which is a correct alarm with well predicted landslide. A pixel is a true-negative (tn) if it is mapped as "stable" both in OL in PL, which corresponds to a well predicted stable area. A pixel is a false-positive (fp) if it is mapped as "unstable" in PL, but is "stable" in OL; that is a false alarm. A pixel is a false-negative (fn) if it is mapped as "stable" in PL, but is "unstable" in OL, that is a missed alarm. The concept of the Receiver Operator Characteristic (ROC, Goodenough et al., 1974) graph is based on the values assumed by tp, fp, tn. ROCs are used to assess the performance of models which provides results assigned to one of two classes. The ROC graph is widely used in many scientific fields such as medicine (Goodenough et al., 1974),

biometrics (Pepe, 2003) and machine learning (Provost and Fawcett, 2001). The ROC graph is a Cartesian plane with the FPR on the x-axis and TPR on the y-axis. FPR is the ratio between false positives and the sum of false positives and true negatives, and TPR is the ratio between true positives and the sum of true positives and false negatives. They are defined in Table 1 and commented on Appendix 1. The performance of a perfect model corresponds to the point P(0,1) on the ROC plane. Points that fall on the bisector (black solid line, on the plots) are associated with models that are considered as random: they predict stable or unstable cells with the same rate.

Eight GOF indices for the quantification of model performances were implemented in the system. Table (1) shows their definition, range, and optimal values. A more comprehensive description of the indices is provided in Appendix 1.

Automatic calibration algorithms implemented in NewAge-JGrass as OMS components can be used in order to tune the model parameters in order to reproduce the actual landslides. This is possible because each model is an OMS component and can be linked to the calibration algorithms as it is, without rewriting or modifying its code. Three calibration algorithms are embedded in the system core: Luca (Hay et al., 2006), a step-wise algorithm based on shuffled complex evolution (Duan et al., 1992), Particle Swarm Optimization (PSO), a genetic model presented in (Kennedy and Eberhart, 1995), and DREAM (Vrugt et al., 2008) an acronym for Differential Evolution Adaptive Metropolis. In the actual configuration we used a Particle Swarm Optimization (PSO) algorithm to estimate optimal values of the model parameters.

During the calibration procedure, the selected algorithm compares the model output in terms of a binary map (stable or unstable pixel) with the actual landslide, thus optimizing a selected objective function (OF). The model parameter set for which the OF assumes its best value is the optimization procedure output. The eight GOF indices presented in Table 1 were used in turn as OFs and, consequently, eight optimal parameters sets were provided as the calibration output (one for each optimised OF). This means that a GOF index selected in Table 1 becomes an OF when it is used as an objective function of the automatic calibration algorithm.

In order to quantitatively analyze the model performances we implemented a three steps verification procedure (3SVP). Firstly, we evaluated the performances of each

OF index for each model. We presented the results in the ROC plane in order to assess what the OF index(es) was (where), whose optimization provided the best model performances. Secondly, we verified wheatear each OF metric had its own information content or wheatear it provided information analogous to other metrics (and thus not essential).

Lastly, for each model, the sensitivity of each optimal parameter set was tested by perturbing optimal parameters and by evaluating their effects on the GOF.

2.4 Site Description

The test site was located in Calabria, Italy, along the Salerno-Reggio Calabria highway between Cosenza and Altilia municipalities, in the southern part of the Crati basin (Figure 2). The mean annual precipitation is about of 1200 mm, distributed over approximately 100 rainy days, with a mean annual temperature of 16 °C. Rainfall peaks occur from October to March, when mass wasting and severe water erosion processes are triggered (Capparelli et al., 2012, Conforti et al., 2011, Iovine et al., 2010).

In the study area the topographic elevation has an average value of around 450 m a.s.l., with a maximum value of 730 m a.s.l. Slopes, computed from the 10 meters resolution digital elevation model, range from 0° to 55°, while the average is about 26°.

The Crati Basin is a Pleistocene-Holocene extensional basin filled by clastic marine and fluvial deposits (Vezzani, 1968; Colella et al., 1987; Fabbricatore et al., 2014). The stratigraphic succession of the Crati Basin can be simply divided into two sedimentary units as suggested by Lanzafame and Tortorici (1986). The first unit is a Lower Pliocene succession of conglomerates and sandstones passing upward into a silty clay (Lanzafame and Tortorici, 1986) second unit. This is a series of clayey deposits grading upward into sandstones and conglomerates which refer to Emilian and Sicilian, respectively (Lanzafame and Tortorici, 1986), as also suggested by data provided by Young and Colella (1988).

In the study area the second unit outcrops. A topsoil of about 1.5 - 2.0 m lies on sandy-gravelly and sandy deposits, which are generally well-stratified. Soils range from Alfisols (i.e. highly mature soils) to Inceptisols and Entisols (i.e. poorly developed soils). Due to the combination of such climatic, geo-structural, and geomorphological features the test site is one of the most landslide prone areas in Calabria (Conforti et al., 2014; Carrara and Merenda,1976; Iovine et al., 2006,).

Mass movements were analyzed from 2006 to 2013 by integrating aerial photography interpretation acquired in 2006, 1:5000 scale topographic maps analysis, and an extensive field survey.

All the data were digitized and stored in a GIS database (Conforti et al., 2014) and the result was the map of occurred landslides, presented in Figure 2,D. Digital elevation model, slope and total contributing area (TCA) maps are presented in Figures 2, A, B, and C respectively. In order to perform model calibration and verification, the dataset of occurred landslides was divided in two parts one used for calibration (located at bottom of Figure 2,D) and one for validation (located in the upper part of Figure 2,D). The landslide inventory map refers only to the initiation area of the landslides. This leads to a fair comparison with the landslide models that provide only the triggering point and does not include a runout model for landslides propagation.

3 RESULTS AND DISCUSSION

The LSA presented in the paper was applied to the Salerno-Reggio Calabria highway, between Cosenza and Altilia (southern Italy). Subsection 3.1 describes the model parameters calibration and the model verification procedure; 3.2 presents the model performance correlation assessment; 3.3 presents the robustness analysis of the GOF indices used; and lastly, 3.4 presents the computation of the susceptibility map.

3.1 Model calibration and verification

The three models presented in Section 2 were used to predict the landslide susceptibility for the study area. Models parameters were optimized using each GOF index presented in Table 1 in order to fit landslides of the calibration group. Table 2 presents the list of parameters that will be optimized, specifying their initial range of variation, and the parameters kept constant during the simulation and their value.

The component PSO provides eigth best parameter sets, one for each optimized GOF indices. Values for each model (M1, M2 and M3) are presented in Table 3. Optimal parameter sets differ slightly among the models and among the optimized GOF indices for a given model. In addition a compensation effect between the parameter values is evident. High values of friction angle are related to low cohesion values; high values of critical rainfall are related to high values of soil resistance parameters. For the model M1, the transmissivity value (74 m^2/d) optimizing ACC is much lower than the transmissivity values obtained by optimizing the other indices (around 140 m^2/d). Similar behavior was observed for the optimal rainfall value which is 148 [mm/d] optimizing ACC, and around 70 [mm/d] optimizing the other indices. For the model M2, the optimal transmissivity and rainfall values optimizing CSI (10 [m^2/d] and 95 [mm/d]), are much lower than the values obtained by optimizing the other indices (around 50 [m^2/d] and 250 [mm/d] in average). For the model M3, on the other hand, optimal parameters present the same order of magnitude for all the optimized indices. This suggests that the variability of the optimal parameter values for models M1 and M2 could be due to compensate the effects of important physical processes neglected by those models.

Executing the models using the eight optimal parameters set, true positive rates and false positive rates are computed by comparing the model output and actual landslides for both the calibration and verification datasets. The results are presented in Table 4, for all three models M1, M2 and M3. These points were reported in the ROC plane to visualize the effects of the optimized objective function on model performances in a unique graph. This procedure was repeated for the three models. ROC planes, considering all the GOF indices and all three models, are included in Appendix 2 both for the calibration and verification period. For models M2 and M3, it is clear that ACC, HSS, and CSI performed the worst. This is also true for model M1, although, unlike M2 and M3, there is no clear separation between the performances provided by ACC, HSS, and CSI and the remaining indices.

Among the results provided in Table 4, we focused on the GOF indices, whose optimization satisfies the condition: FPR<0.4 and TPR>0.7. This choice was made in order to focus comments on the results exclusively for the GOF indices which provide acceptable model results and in order to heighten the readability of graphs.

Figure 3 presents three ROC planes, one for each model, with the optimized GOF indices that provide FPR<0.4 and TPR>0.7. The results presented in Figure 3 and Table 4 show that: i) the optimization of AI, D2PC, SI and TSS achieves the best model performance in the ROC plane, which is verified for all three models; ii) performances increase as model complexity increases: moving from M1 to M3 points in the ROC plane approaches the perfect point (TPR=1, FPR=0); iii) by increasing the model complexity, good model results are achieved, not only in the calibration but also in the validation dataset. In fact, moving from M1 to M2 soil cohesion and soil properties were considered, and moving from M2 to M3 rainfall of a finite duration was used.

The first step of the 3SVP procedure highlights that the optimization of AI, D2PC, SI, and TSS provides the best performances irrespectively of the model used.

Finally, it is important to consider the limitations of the models used for the current applications. Models M1 and M2 are not able to mimic the transient nature of precipitation and infiltration processes, and only M3 is able to account for the combined effect of storm duration and intensity in the triggering mechanism. In addition, in this study we neglected effects such as spatial rainfall variability, roads, and other engineering works.

3.2 Correlations assessment of the models performances

The second step in the procedure is to verify the information content of each optimized OF, checking whether it is the same as other metrics or it is particular feature of the optimized OF.

Executing a model using one of the eight parameters set (assuming, for example, the one obtained by optimizing CSI) enables all the remaining GOF indices to be computed, which we indicate as CSI_{CSI}, ACC_{CSI}, HSS_{CSI}, TSS_{CSI}, AI_{CSI}, SI_{CSI}, $D2PC_{CSI}$, ESI_{CSI}, both for calibration and for verification dataset. Let us denote this vector with the name $\boldsymbol{MP_{CSI}}$: the model performance (***MP***) vector computed using the parameter set that optimizes CSI. $\boldsymbol{MP_{CSI}}$ has 16 elements, 8 for the calibration and 8 for the validation dataset. Repeating the same procedure for all eight GOF indices it gives: $\boldsymbol{MP_{ACC}}$, $\boldsymbol{MP_{ESI}}$, $\boldsymbol{MP_{SI}}$, $\boldsymbol{MP_{D2PC}}$, $\boldsymbol{MP_{TSS}}$, $\boldsymbol{MP_{AI}}$, $\boldsymbol{MP_{HS}}$. Figure 4 presents the correlation plots (Murdoch and Chow, 1996) between all ***MP*** vectors, for each model

M1, M2 or M3. The matrix is symmetric with an ellipse at the intersection of row i and column j. The color is the absolute value of the correlation coefficient between the $\boldsymbol{MP_i}$ and $\boldsymbol{MP_j}$ vectors. The eccentricity of the ellipse is scaled according to the correlation value: the more prominent it is, the less correlated are the vectors. If the ellipse leans towards the right, the correlation is positive, if it leans to the left, it is negative.

All indices present a positive correlation with each other, irrespectively of the model used. In addition, strong correlations between the ***MP*** vectors of AI, D2PC, SI, and TSS are evident in Figure 4. This confirms that an optimization of AI, D2PC, SI, and TSS provides similar model performances, irrespectively of the model used. On the other hand, the remaining GOF indices give quite different information from the previous four indices, however their performance was worse in the first step of the analysis. Thus in the case study, using one of the four best GOFs is sufficient for the parameter estimation.

3.3 Models sensitivity assessment

In this step we focused on models M2 and M3 and performed a parameter sensitivity analysis. Let us consider model M2 and the optimal parameter set computed by optimizing the Critical Success Index (CSI). Also, considering the cohesion model parameter, the procedure evolves according to the following steps:

- The starting parameter values are the optimal values derived from the optimization of the CSI index;
- All the parameters except the analyzed parameter (cohesion) were kept constant and equal to the optimal parameter set;
- 1000 random values of the analyzed parameter (cohesion) were selected from a uniform distribution with the lower and upper bound defined in Table 1. With this procedure 1000 model parameter sets were defined and used to execute the model.
- 1000 values of the selected GOF index (CSI), computed by comparing model outputs with the measured data, were used to compute a boxplot of the parameter C and optimized index CSI.

The procedure was repeated for each parameter and for each optimized index. Results are presented in Figures 5 and 6 for models M2 and M3 respectively.

Each column in the figures represents one optimized index and has a number of boxplots equal to the number of model parameters (5 for M2 and 6 for M3). Each boxplot represents the range of variation of the optimized index due to a particular change in the model parameters. The narrower the boxplot for a given optimized index, the less sensitive the model is to that parameter. For both M2 and M3, the parameter set obtained by optimizing AI and SI shows the least sensitive behavior for almost all the parameters. In this case a model parameter perturbation has little impact on the model's performances. However, the models with parameters obtained by optimizing ACC, TSS, and D2PC are the most sensitive to the parameter variations and this is reflected in much more evident changes in model performances. Finally, it is important to consider that the methodology used for evaluating the parameter sensitivity is based on changing the parameters one-at-time. Although this procedure facilitates an inter-comparison of the results (because the parameter sensitivity is computed with reference to the optimal parameter set), it is does not take into account simultaneous variations or interactions between parameters.

3.4 Models selections and susceptibility maps

The selection of the most appropriate model for computing landslide susceptibility maps is based on what we learn from the previous steps. In the first step we learn that i) the optimization of AI, D2PC, SI and TSS outperforms the remaining indices and ii) models M2 and M3 provide more accurate results than M1. The second step suggests that overall the model results obtained by optimizing AI, D2PC, SI and TSS are similar each other. Lastly, the third step shows that the model performance derived from the optimization of AI and SI is less sensitive to input variations than D2PC and TSS. This could be due to the formulation of AI and SI which gives much more weight to the true negative compared to D2PC and TSS.

For our application, the model M3 with parameters obtained by optimizing D2PC was the most sensitive to the parameter variation avoiding, an “insensitive” or flat response by changing the parameters values. A more sensitive couple model-

optimal parameter set will in fact accommodate any parameters, input data, or measured data variations responding to these changes with a variation in model performance.
We thus used the combination of model M3 with parameters obtained by optimizing D2PC in order to compute the final susceptibility maps in Figure 7. Categories of landslide susceptibility from classes 1 to 5 are assigned from low to high according to FS values (e.g. Huang et al., 2007): Class 1 ($FS \leq 1.0$), Class 2 ($1.0<FS<1.2$), Class 3 ($1.2<FS<1.5$), Class 4 ($1.5<FS<2.0$), Class 5 ($FS \geq 2$).

4 Conclusions

We have presented a procedure to quantitatively calibrate, evaluate, and compare the performances of environmental models. The procedure was applied for the analysis of three landslides susceptibility models. It is made up of three steps: i) model parameters calibration, optimizing different GOF indices and models evaluation in the ROC plane; ii) computation of the degree of similarities between different model performances obtained by optimizing all the considered GOF indices; iii) evaluation of model sensitivity to parameter variations. The first step identifies the more appropriate OFs for the model parameter optimization. The second step verifies the information content of each optimized OF, checking whether it is analogous to other metrics or peculiar to the optimized OF. Finally the last step quantifies the relative influence of each model parameter on the model performance.
The procedure was conceived as a model configuration of the hydrological system NewAge-JGrass; it integrates: i) three simplified physically based landslides susceptibility models; ii) a package for model evaluations based on pixel-by-pixel comparison of modeled and actual landslides maps; iii) models parameters calibration algorithms, and iv) the integration with the uDig open-source geographic information system for model input-output map management. The system is open-source and available at (https://github.com/formeppe). It is integrated according to the Object Modeling System standards which enables the user to easily integrate a generic landslide susceptibility model and use the complete framework presented in the paper, thus avoiding having to rewrite programming code.

The procedure was applied in a test case on the Salerno-Reggio Calabria highway and led to the following conclusions: 1) the OFs AI, D2PC, SI, and TSS coupled with the models M2 and M3 provided the best performances among the eights metrics used in the calibration; 2) the four selected OFs provided quite similar model performances in terms of MP vectors, i.e. one of them would be sufficient for the model application; 3) M3 showed the best performance by optimizing the D2PC index. In fact M3 responded to parameter variations with changes in model performances.

In our application effective precipitation was calibrated because we were performing a landslide susceptibility analysis and it was useful for demonstrating the method. However, we are aware that for operational landslide early warning systems, rainfall constitutes a fundamental input of the predictive process. In addition, the analysis would profit from data on the rainfall that triggered the landslides, however such data are currently not available for the study area.

We believe that our system would be useful for decision makers who deal with risk management assessments. It could be improved by adding new landslide susceptibility models or different types of model selection procedures.

ACKNOWLEDGMENTS

This research was funded by the PON Project No. 01_01503 “Integrated Systems for Hydrogeological Risk Monitoring, Early Warning and Mitigation Along the Main Lifelines”, CUP B31H11000370005, within the framework of the National Operational Program for "Research and Competitiveness" 2007-2013. The authors would like to acknowledge the editor and the three reviewers (Prof. M. Mergili and two unknown reviewers) for providing insightful comments and improving the quality of the paper.

Acronyms table

3SVP	Three steps verification procedure
AI	Average Index
CSI	Critical success index
D2PC	Distance to perfect classification
ESI	Equitable success index
fn	False negative
fp	False positive
FPR	False positive rate
FS	Factor of safety
GIS	Geographic informatic system
GOF	Goodness of fit indices
HSS	Heidke skill score
LSA	Landslide susceptibility analysis
M1	Model for landslide susceptibility analysis proposed in Montgomery and Dietrich, 1994
M2	Model for landslide susceptibility analysis proposed in Park et al., 2013
M3	Model for landslide susceptibility analysis proposed in Rosso et al., 2006
MP	Model performances vector
OF	Objective function
OL	Observed landslide map
OMS	Object modeling system
PL	Predicted landslide map
PSO	Particle Swarm optimization
ROC	Receiver operating characteristic
SI	Success index
TCA	Total contributing area
tn	True negative
tp	True positive
TPR	True positive rate
TSS	True Skill Statistic

REFERENCES

Abera W., A. Antonello, S. Franceschi, G. Formetta, R Rigon , "The uDig Spatial Toolbox for hydro-geomorphic analysis" in Geomorphological Techniques, v. 4, n. 1 (2014), p. 1-19. - URL: http://www.geomorphology.org.uk/sites/default/files/geom_tech_chapters/2.4.1_GISToolbox.pdf

Beguería, S. (2006). Validation and evaluation of predictive models in hazard assessment and risk management. Natural Hazards, 37(3), 315-329.

Bennett ND, Croke BF, Guariso G, Guillaume JH, Hamilton SH, Jakeman AJ, Marsili-Libelli S, Newham LT, Norton JP, Perrin C, Pierce SA. Characterising performance of environmental models. Environmental Modelling & Software. 2013 Feb 28;40:1-20.

Borga, M., Dalla Fontana, G., & Cazorzi, F. (2002). Analysis of topographic and climatic control on rainfall-triggered shallow landsliding using a quasi-dynamic wetness index. Journal of Hydrology, 268(1), 56-71.

Brabb, E.E., (1984). Innovative approaches to landslide hazard and risk mapping, Proceedings of the 4th International Symposium on Landslides, 16–21 September, Toronto, Ontario, Canada (Canadian Geotechnical Society, Toronto, Ontario, Canada), 1:307–324

Brenning, A. "Spatial prediction models for landslide hazards: review, comparison and evaluation." Natural Hazards and Earth System Science 5, no. 6 (2005): 853-862.

Capparelli, G., & Versace, P. (2011). FLaIR and SUSHI: two mathematical models for early warning of landslides induced by rainfall. Landslides, 8(1), 67-79.

Capparelli G, Iaquinta P, Iovine GGR, Terranova OG, Versace P. Modelling the rainfall-induced mobilization of a large slope movement in northern Calabria. Natural Hazards 2012 ;61:247–256.

Carrara, A., Merenda, L., 1976. Landslide inventory in Northern Calabria, Southern Italy. Geological Society of America Bulletin 87, 1153–1162

Casadei, M., Dietrich, W. E., & Miller, N. L. (2003). Testing a model for predicting the timing and location of shallow landslide initiation in soil‐mantled landscapes. Earth Surface Processes and Landforms, 28(9), 925-950.

Cascini, L., Bonnard, C., Corominas, J., Jibson, R., & Montero-Olarte, J. (2005). Landslide hazard and risk zoning for urban planning and development. Landslide Risk Management. Taylor and Francis, London, 199-235.

Catani, F., Casagli, N., Ermini, L., Righini, G., & Menduni, G. (2005). Landslide hazard and risk mapping at catchment scale in the Arno River basin. Landslides, 2(4), 329-342.

Chung C-JF, Fabbri AG and van Westen CJ (1995) Multivariate regression analysis for landslide hazard zonation. Carrara A and Guzzetti F (Eds.) Geographical Information Systems in assessing natural hazards. Dordrecht, Kluwer Academic Publishers. 5:107-34

Colella A, De Boer PL, Nio SD. Sedimentology of a marine intermontane Pleistocene Gilbert-type fan-delta complex in the Crati Basin, Calabria, southern Italy. Sedimentology 1987;34:721–736.

Conforti, M., Pascale, S., Robustelli, G., & Sdao, F. (2014). Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). Catena, 113, 236-250.

Conforti M, Aucelli PPC, Robustelli G, Scarciglia F. Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo Stream catchment (Northern Calabria, Italy). Natural Hazards 2011;56:881–898.

Corominas J, Van Westen C, Frattini P, Cascini L, Malet JP, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, Pitilakis K. Recommendations for the quantitative analysis of landslide risk. Bulletin of engineering geology and the environment. 2014 May 1;73(2):209-63.

Dietrich, W. E., Bellugi, D. and Real De Asua, R. (2001) Validation of the Shallow Landslide Model, SHALSTAB, for Forest Management, in Land Use and Watersheds: Human Influence on Hydrology and Geomorphology in Urban and Forest Areas (eds M. S. Wigmosta and S. J. Burges), American Geophysical Union, Washington, D. C.. doi: 10.1029/WS002p0195

David, O., Ascough II, J. C., Lloyd, W., Green, T. R., Rojas, K. W., Leavesley, G. H., & Ahuja, L. R. (2013). A software engineering perspective on environmental modeling framework design: The Object Modeling System. Environmental Modelling & Software, 39, 201-213.

Duan, Q., Sorooshian S., and Gupta V(1992): Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research 28.4 (1992): 1015-1031.

Fabbricatore D, Robustelli G, Muto F. Facies analysis and depositional architecture of shelf-type deltas in the Crati Basin (Calabrian Arc, south Italy). Boll. Soc. Geol. It. 2014;133(1):131-148.

Formetta, G., Mantilla, R., Franceschi, S., Antonello, A., & Rigon, R. (2011). The JGrass-NewAge system for forecasting and managing the hydrological budgets at the basin scale: models of flow generation and propagation/routing. Geoscientific Model Development, 4(4), 943-955.

Formetta, G., Antonello, A., Franceschi, S., David, O., & Rigon, R. (2014). Hydrological modelling with components: A GIS-based open-source framework. Environmental Modelling & Software, 55, 190-200.

Formetta, G., Capparelli, G., Rigon, R., and Versace, P.: Physically based landslide susceptibility models with different degree of complexity: calibration and verification. International Environmental Modelling and Software Society (iEMSs). 7th Intl. Congress on Env. Modelling and Software, San Diego, CA, June 15-19, USA, Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.), 2014. http://www.iemss.org/sites/iemss2014/papers/iemss2014_submission_157.pdf

G. Formetta, G. Capparelli, and P. Versace , Modelling rainfall induced shallow landslides in the Landslide Early Warning Integrated System project Slopes and Geohazards. January 2015, 1747-1752. Available at: http://www.icevirtuallibrary.com/doi/abs/10.1680/ecsmge.60678.vol4.260

Formetta, G., Simoni, S., Godt, J. W., Lu, N., & Rigon, R. (2016). Geomorphological control on variably saturated hillslope hydrology and slope instability. Water Resources Research.

Frattini, P., Crosta, G., & Carrara, A. (2010). Techniques for evaluating the performance of landslide susceptibility models. Engineering geology, 111(1), 62-72.

Guzzetti, Fausto, Alberto Carrara, Mauro Cardinali, and Paola Reichenbach. "Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy." Geomorphology 31, no. 1 (1999): 181-216.

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

Glade, T., & Crozier, M. J. (2005). A review of scale dependency in landslide hazard and risk analysis. Landslide hazard and risk, Vol. 3, 75-138.

Goodenough, D.J., Rossmann, K., Lusted, L.B., 1974. Radiographic applications of receiver operating characteristic (ROC) analysis. Radiology 110, 89–95.

Grahm J (1984) Methods of slope stability analysis. In: Brunsden D, Prior DB (eds) Slope instability. Wiley, New York, pp 171–215

Hay, L.E., G.H. Leavesley, M.P. Clark, S.L. Markstrom, R.J. Viger, and M. Umemoto (2006). Step-Wise, Multiple-Objective Calibration of a Hydrologic Model for a Snowmelt-Dominated Basin. Journal of the American Water Resources Association 42:877-890, 2006

Huang, J. C., Kao, S. J., Hsu, M. L., & Liu, Y. A. (2007). Influence of Specific Contributing Area algorithms on slope failure prediction in landslide modeling. Natural Hazards and Earth System Science, 7(6), 781-792.

Lanzafame G, Tortorici L. La tettonica recente del Fiume Crati (Calabria). Geografia Fisica e Dinamica Quaternaria 1984; 4:11-21.

Iovine, G., Petrucci, O., Rizzo, V., Tansi, C., 2006. The March 7th 2005 Cavallerizzo (Cerzeto) landslide in Calabria—Southern Italy. Engineering geology for tomorrow's cities—the 10th IAEG congress, Nottingham (UK), The Geological Society of London, Paper number 785.

Lee, S., Choi, J., & Min, K. (2002). Landslide susceptibility analysis and verification using the Bayesian probability model. Environmental Geology, 43(1-2), 120-131.

Kennedy, J., and Eberhart R.(1995): Particle swarm optimization. Neural Networks, 1995. Proceedings., IEEE International Conference on. Vol. 4. Perth, WA. IEEE, 1995.

Iovine GGR, Lollino P, Gariano SL, Terranova OG. Coupling limit equilibrium analyses and real-time monitoring to refine a landslide surveillance system in Calabria (southern Italy). Natural Hazards and Earth System Sciences 2010; 10:2341–2354.

Jolliffe, I. T., & Stephenson, D. B. (Eds.). (2012). Forecast verification: a practitioner's guide in atmospheric science. University of Exeter, UK.

John Wiley & Sons.

Lu, N., and J. Godt (2008), Infinite slope stability under steady unsaturated seepage conditions, Water Resour. Res., 44, W11404, doi:10.1029/2008WR006976.

Lu, N., & Godt, J. W. (2013). Hillslope hydrology and stability. Cambridge University Press.

Milledge, D. G., Bellugi, D., McKean, J. A., Densmore, A. L., & Dietrich, W. E. (2014). A multidimensional stability model for predicting shallow landslide size and shape across landscapes. Journal of Geophysical Research: Earth Surface, 119(11), 2481-2504.

Montgomery, D. R., & Dietrich, W. E. (1994). A physically based model for the topographic control on shallow landsliding. Water resources research, 30(4), 1153-1171.

Murdoch, D. J., & Chow, E. D. (1996). A graphical display of large correlation matrices. The American Statistician, 50(2), 178-180.

Naranjo, J.L., van Westen, C.J. and Soeters, R. (1994) Evaluating the use of training areas in bivariate statistical landslide hazard analysis: a case study in Colombia. ITC Journal, 3:292-300.

Pepe, M.S., 2003. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press, New York.

Park, N. W. (2011). Application of Dempster-Shafer theory of evidence to GIS-based landslide susceptibility analysis. Environmental Earth Sciences, 62(2), 367-376.

Park, H. J., Lee, J. H., & Woo, I. (2013). Assessment of rainfall-induced shallow landslide susceptibility using a GIS-based probabilistic approach. Engineering Geology, 161, 1-15.

Pradhan, B. (2011). An assessment of the use of an advanced neural network model with five different training strategies for the preparation of landslide susceptibility maps. Journal of Data Science, 9(1), 65-81.

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Computers & Geosciences, 51, 350-365.

Provost, F., Fawcett, T., 2001. Robust classification for imprecise environments. Machine Learning 42 (3), 203–231.

Rosso, R., M. C. Rulli, and G. Vannucchi (2006), A physically based model for the hydrologic control on shallow landsliding, Water Resour. Res., 42, W06410, doi:10.1029/2005WR004369.

Sidle, R. C., & Ochiai, H. (2006). Landslides: processes, prediction, and land use (Vol. 18). Washington, DC 20009, USA. American Geophysical Union.

Simoni, S., Zanotti, F., Bertoldi, G., and Rigon, R. (2008): Modeling the probability of occurrence of shallow landslides and channelized debris flows using GEOtop-FS, Hydrol. Process., 22, 532{545,

Vezzani L. I terreni plio-pleistocenici del basso Crati (Cosenza). Atti dell'Accademia Gioenia di Scienze Naturali di Catania 6:28–84; 1968.

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44, W00B09, doi:10.1029/2007WR006720.

Young J, Colella A. Calcarenous nannofossils from the Crati Basin. In: Colella A. (ed.), Fan Deltas-Excursion Guidebook. Università della Calabria, Cosenza, Italy. 79-96; 1988.

Table 1: Indices of goodness of fit for comparison between actual and predicted landslide.

Name	Definition	Range	Optimal value
Critical success index (CSI)	$\mathrm{CSI}=\frac{\mathrm{tp}}{\mathrm{tp+fp+fn}}$	[0 ,1]	1.0
Equitable success index (ESI)	$ESI=\frac{\mathrm{tp-R}}{\mathrm{tp+fp+fn-R}}$ $R=\frac{(tp+fn)\cdot(tp+fp)}{tp+fn+fp+tn}$	[-1/3,1]	1.0
Success Index (SI)	$\mathrm{SI}=\frac{1}{2}\cdot\left(\frac{tp}{tp+fn}+\frac{tn}{fp+tn}\right)$	[0 ,1]	1.0
Distance to perfect classification (D2PC)	$D2PC=\sqrt{(1-TPR)^2+FPR^2}$ $\mathrm{TPR}=\frac{\mathrm{tp}}{\mathrm{tp+fn}}$ $\mathrm{FPR}=\frac{\mathrm{fp}}{\mathrm{fp+tn}}$	[0,1]	0.0
Average Index (AI)	$AI=\frac{1}{4}\left(\frac{tp}{tp+fn}+\frac{tp}{tp+fp}+\frac{tn}{fp+tn}+\frac{tn}{fn+tn}\right)$	[0,1]	1.0
True skill statistic (TSS)	$TSS=\frac{(tp\cdot tn)-(fp\cdot fn)}{(tp+fn)\cdot(fp+tn)}$	[-1,1]	1.0
Heidke skill score (HSS)	$HSS=\frac{2\cdot(tp\cdot tn)-(fp\cdot fn)}{(tp+fn)\cdot(fn+tn)+(tp+fp)\cdot(fp+tn)}$	[-∞, 1]	1.0
Accuracy (ACC)	$ACC=\frac{(\mathrm{tp}+\mathrm{tn})}{(\mathrm{tp}+\mathrm{fn}+\mathrm{fp}+\mathrm{tn})}$	[0,1]	1.0

Table 2: Optimised models' parameters values

Model Parameters	**Constant Value**	**Range value**
Soil Depth [m]	-	[0.8; 5.0]
Transmissivity [m^2/d]	-	[10; 150]
Soil/water density ratio	-	[1.8; 2.8]
Friction Angle [°]	-	[11; 40]
Rainfall [mm/d]	-	[50; 300]
Soil Cohesion [kPa]	-	[0; 50]
Degree Of Saturation [-]	0.5	-
Soil Porosity [-]	0.5	-
Rainfall Duration [d]	-	[0.1; 3.0]

Table 3: Optimal parameter sets output of the optimization procedure of each GOF indices in turn. Results are presented for each model (M1, M2 and M3).

Model: M1								
Optimised Index	AI	HSS	TSS	D2PC	SI	ESI	CSI	ACC
Soil Depth [m]	1.32	1.85	1.44	2.80	1.36	2.62	2.42	2.01
Transmissivity [m^2/d]	140.24	146.31	142.68	137.10	147.69	144.66	136.73	74.74
Soil/water density ratio [-]	2.61	2.56	2.77	2.71	2.78	2.79	2.63	2.72
Friction Angle [°]	24.20	32.40	22.50	23.10	22.40	29.50	29.50	38.30
Rainfall [mm/d]	85.38	53.30	71.36	50.00	52.69	69.19	61.35	141.80

Model: M2								
Optimised Index	AI	HSS	TSS	D2PC	SI	ESI	CSI	ACC
Transmissivity [m^2/d]	65.43	33.22	80.45	38.22	84.54	33.24	10.70	55.76
Cohesion [kPa]	25.17	49.63	49.42	16.94	30.01	41.24	44.58	46.85
Friction Angle [°]	29.51	38.38	20.01	32.30	24.57	33.78	35.68	34.96
Rainfall [mm/d]	236.14	293.44	270.42	153.61	294.70	298.44	95.35	299.01
Soil/water density ratio [-]	2.11	2.40	2.06	2.44	2.77	2.17	2.55	2.19
Soil Depth [m]	2.35	1.68	2.38	2.44	2.74	1.12	1.37	1.12

Model: M3								
Optimised Index	AI	HSS	TSS	D2PC	SI	ESI	CSI	ACC
Transmissivity [m^2/d]	30.95	26.55	47.03	36.31	57.28	25.84	31.60	48.71
Cohesion [kPa]	36.88	44.33	28.51	31.60	45.46	41.80	32.05	37.09
Friction Angle [°]	19.55	36.44	27.80	29.70	21.46	33.27	36.47	38.50
Rainfall [mm/d]	248.77	230.08	258.82	201.71	299.90	291.32	273.03	193.02
Soil/water density ratio [-]	2.40	2.57	2.08	2.80	2.65	2.63	2.61	2.44
Soil Depth [m]	1.84	1.42	2.23	2.92	2.85	1.17	1.13	1.15
Rainfall Duration [d]	0.12	1.78	1.24	1.96	1.24	0.39	1.30	1.98

Table 4: Results in term of true-positive rate (TPR) and false-positive rate (FPR), for each model (M1, M2 and M3), for each optimised GOF index and for both calibration (CAL) and verification (VAL) dataset. In bold are shown the rows for which the condition FPR<0.4 and TPR>0.7 is verified.

		MODEL: M1		MODEL: M2		MODEL: M3	
Period	Optim. Index	FPR	TPR	FPR	TPR	FPR	TPR
CAL	ACC	0.04	0.12	0.03	0.12	0.03	0.13
CAL	**AI**	**0.29**	**0.70**	**0.35**	**0.79**	**0.38**	**0.82**
CAL	CSI	0.17	0.48	0.10	0.36	0.09	0.32
CAL	**D2PC**	**0.32**	**0.72**	**0.32**	**0.76**	**0.32**	**0.75**
CAL	ESI	0.17	0.48	0.43	0.82	0.09	0.36
CAL	HSS	0.12	0.35	0.09	0.35	0.09	0.35
CAL	**SI**	**0.34**	**0.74**	**0.39**	**0.85**	**0.39**	**0.86**
CAL	**TSS**	**0.34**	**0.73**	**0.39**	**0.83**	**0.37**	**0.82**
VAL	ACC	0.05	0.12	0.03	0.12	0.03	0.10
VAL	AI	0.26	0.56	0.31	0.69	**0.34**	**0.72**
VAL	CSI	0.17	0.39	0.09	0.31	0.08	0.29
VAL	D2PC	0.29	0.59	0.28	0.67	0.28	0.66
VAL	ESI	0.17	0.39	0.41	0.76	0.09	0.30
VAL	HSS	0.12	0.30	0.09	0.30	0.09	0.30
VAL	SI	0.30	0.61	**0.37**	**0.75**	**0.39**	**0.76**
VAL	TSS	0.30	0.62	**0.35**	**0.74**	**0.34**	**0.71**

Figure 1: Integration of the Landslide susceptibility analysis system in NewAge-JGrass hydrological model.

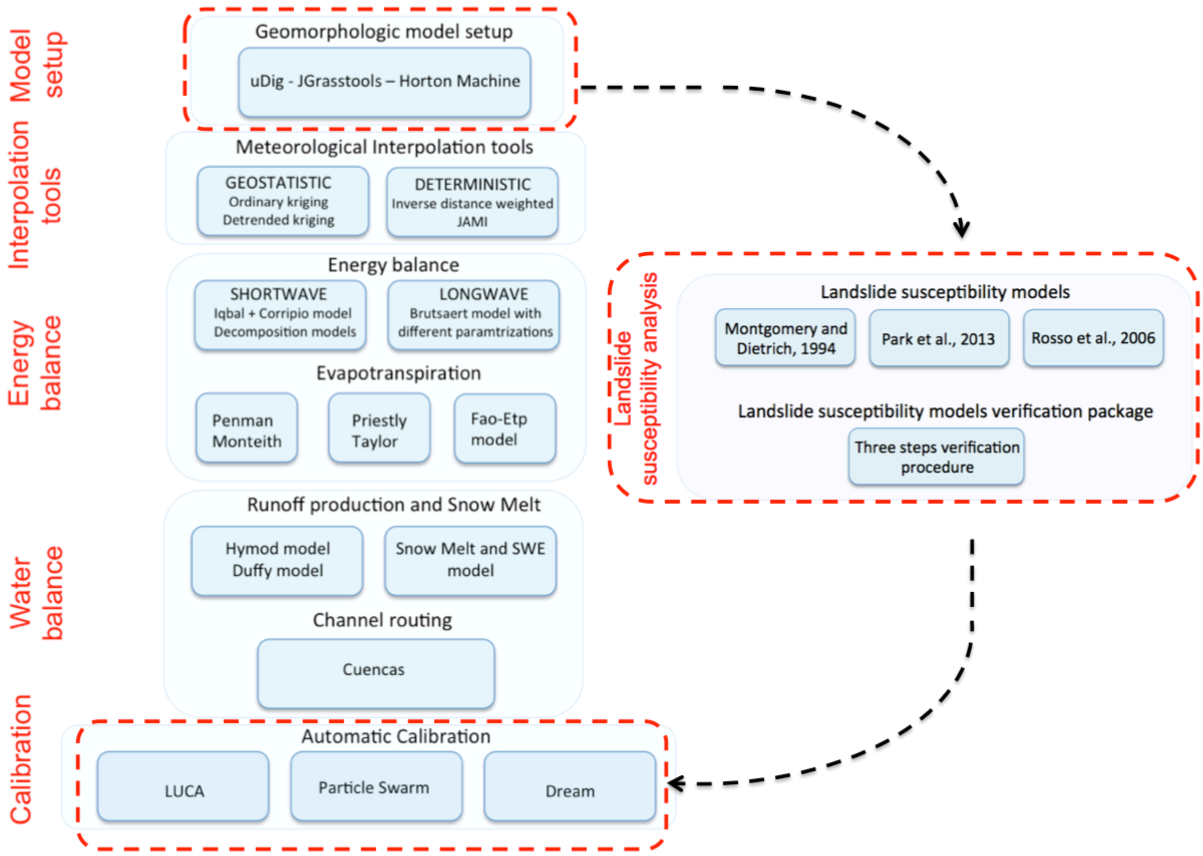
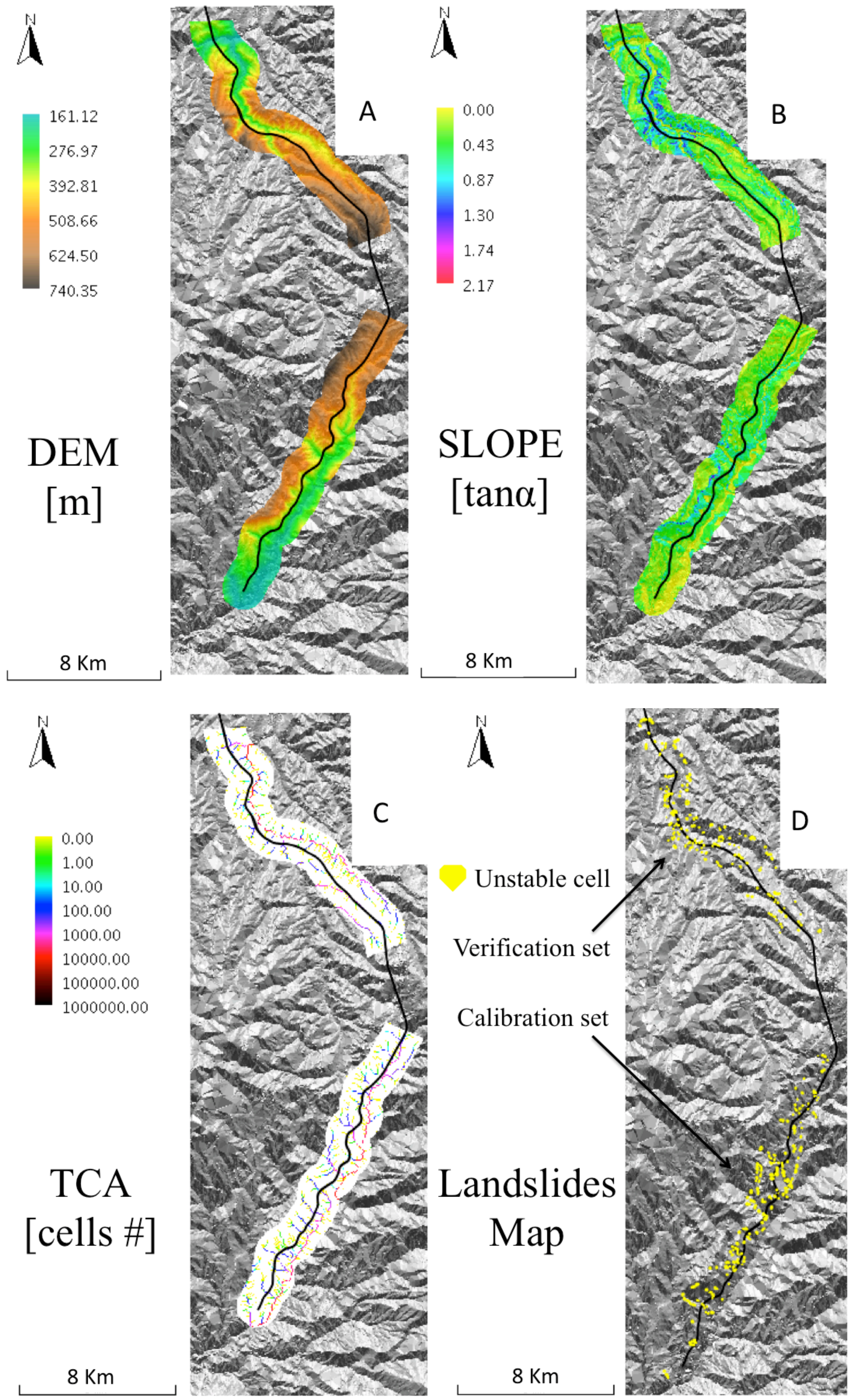

Model setup
Interpolation tools
Energy balance
Water balance
Calibration
Geomorphologic model setup
uDig - JGrasstools – Horton Machine
Meteorological Interpolation tools
GEOSTATISTIC
Ordinary kriging
Detrended kriging
DETERMINISTIC
Inverse distance weighted
JAMI
Energy balance
SHORTWAVE
Iqbal + Corripio model
Decomposition models
LONGWAVE
Brutsaert model with
different paramtrizations
Evapotranspiration
Penman Monteith
Priestly Taylor
Fao-Etp model
Runoff production and Snow Melt
Hymod model
Duffy model
Snow Melt and SWE model
Channel routing
Cuencas
Automatic Calibration
LUCA
Particle Swarm
Dream
Landslide susceptibility analysis
Landslide susceptibility models
Montgomery and Dietrich, 1994
Park et al., 2013
Rosso et al., 2006
Landslide susceptibility models verification package
Three steps verification procedure

Figure 2: Test site. A) Digital elevation model (DEM) [m], B) slope [-] expressed as tangent of the angle, C) total contributing area (TCA) expressed as number of draining cells and D) Map of actual landslides.



DEM
[m]
161.12
276.97
392.81
508.66
624.50
740.35
A
8 Km
SLOPE
[tanα]
0.00
0.43
0.87
1.30
1.74
2.17
B
8 Km
TCA
[cells #]
0.00
1.00
10.00
100.00
1000.00
10000.00
100000.00
1000000.00
C
8 Km
Unstable cell
Verification set
Calibration set
Landslides
Map
D
8 Km

Figure 3: Models' performances results in the ROC plane for M1, M2 and M3. Only GOF indices whose optimization provides FPR<0.4 and TPR>0.7 were reported.

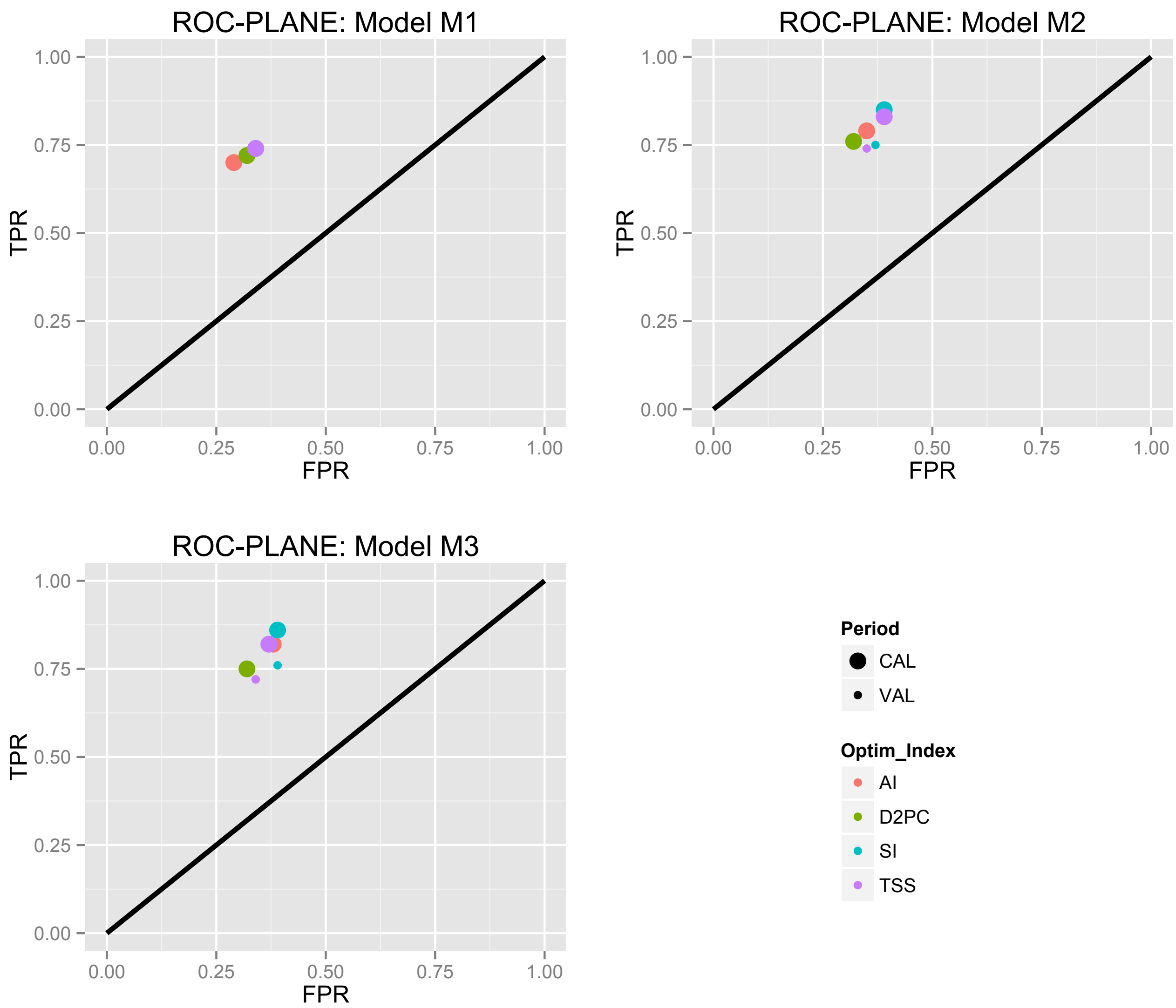


Figure 4: Correlation plot between models' performance (MP) vector computed by optimizing all GOF indices in turn. Results are reported for each model: M1, M2 and M3.

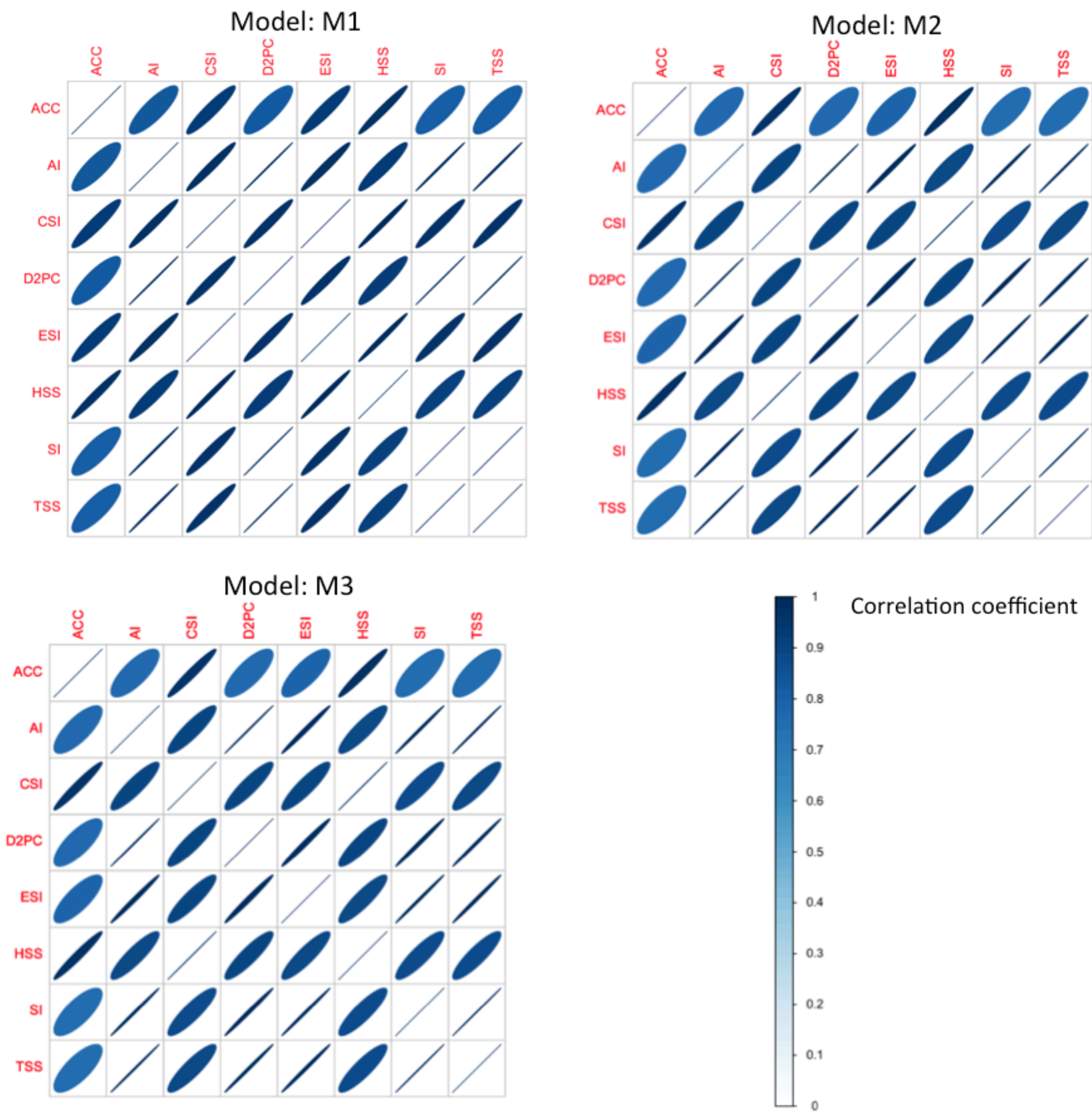


Figure 5: Model M2 parameters sensitivity analysis.

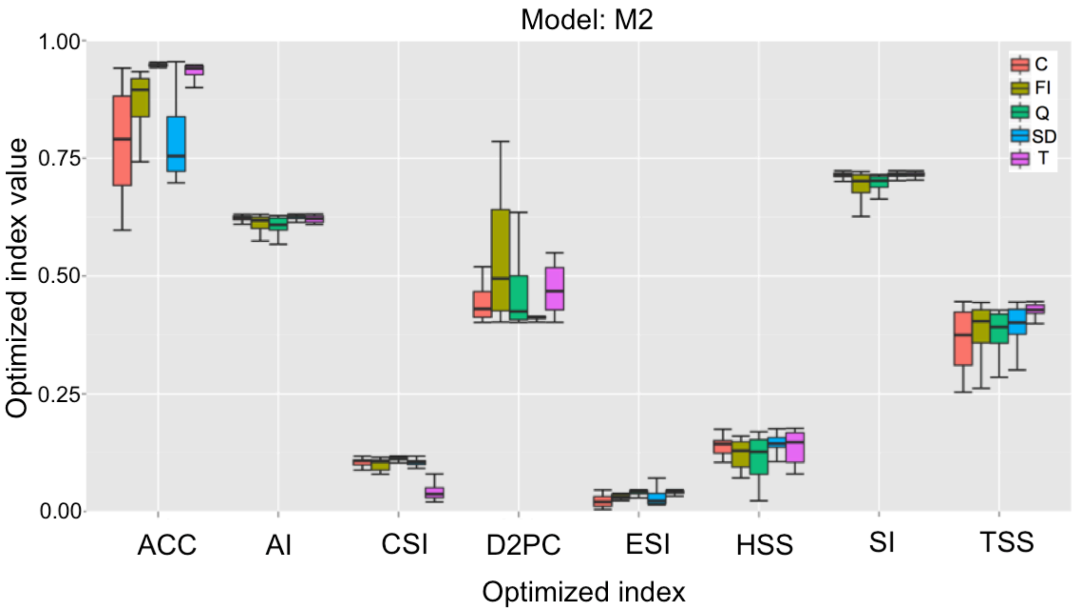


Figure 6: Model M3 parameters sensitivity analysis.

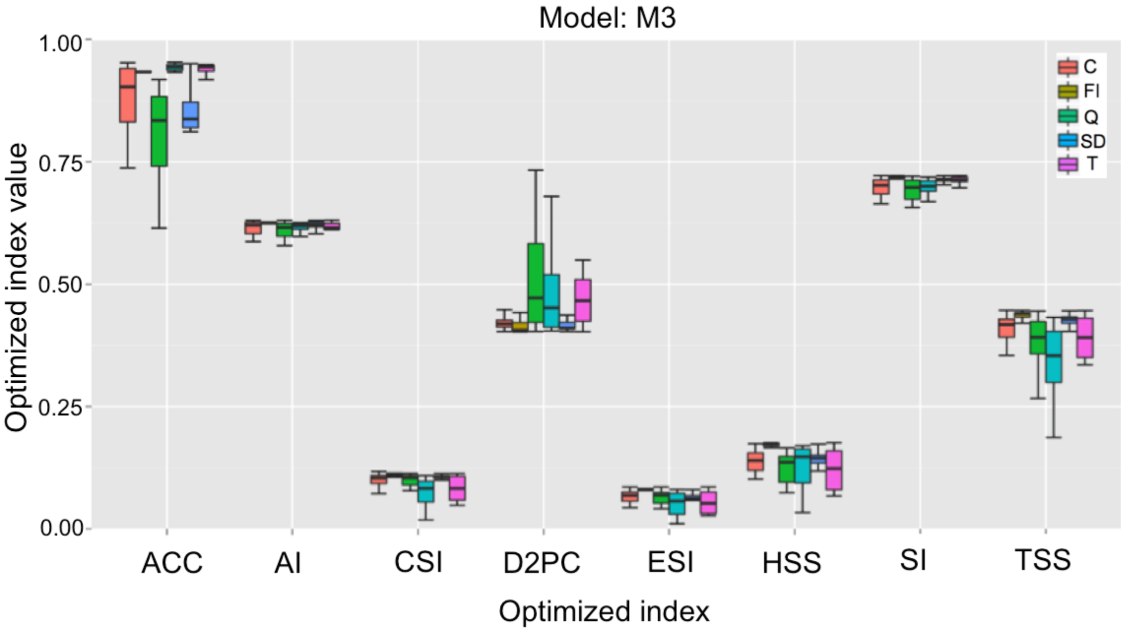
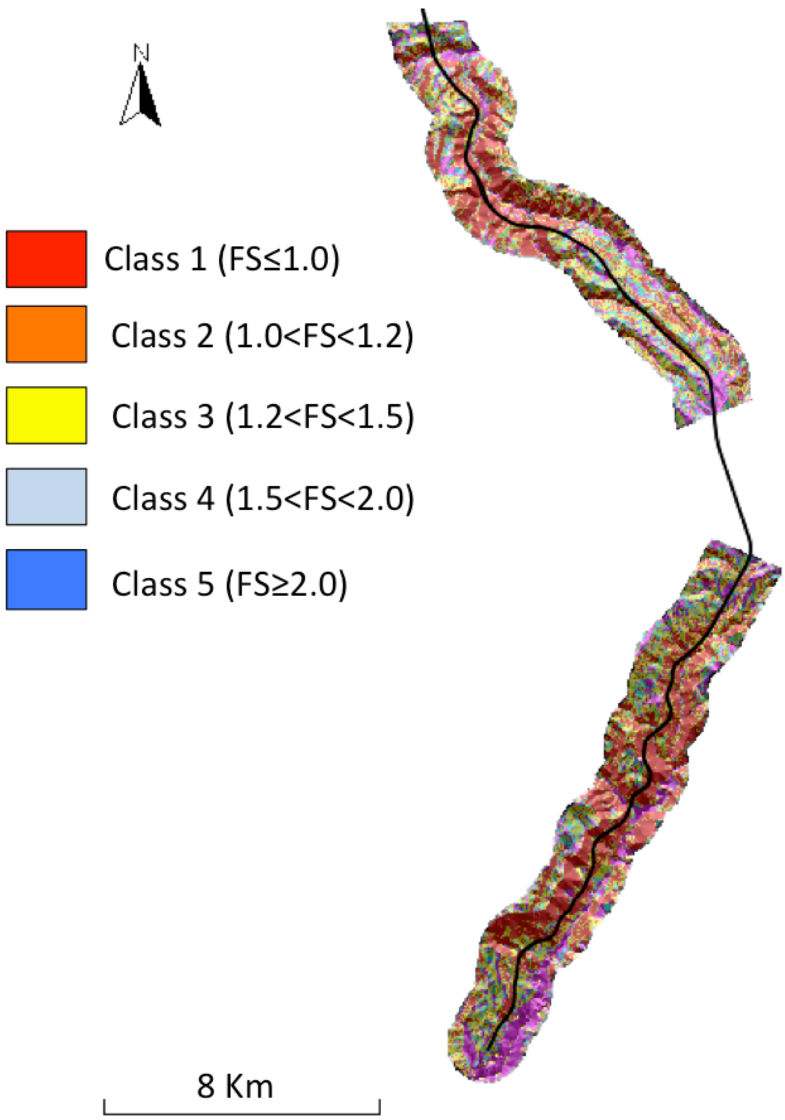


Figure 7: Landslide susceptibility maps using model M3 and parameter set obtained by optimising D2PC.



Appendix 1

1.2 Critical success index (CSI)

CSI, eq. (2), is the number of correct detected lindslide pixels (tp), divided by the sum of tp, fn and fp. CSI is also named threat score. It range between 0 and 1 and its best value is 1. It penalizes both fn and fp.

$$\mathrm{CSI}=\frac{\mathrm{tp}}{\mathrm{tp+fp+fn}} \quad (2)$$

1.3 Equitable success index (ESI)

ESI, eq. (3), contrarily to CSI, is able to take into account the true positives associated with random chance (R). ESI ranges between -1/3 and 1. Value 1 indicates perfect score.

$$E\mathrm{SI}=\frac{\mathrm{tp\text{-}R}}{\mathrm{tp+fp+fn\text{-}R}} \quad 3)$$

$$R=\frac{(tp+fn)\cdot(tp+fp)}{tp+fn+fp+tn} \quad (4)$$

1.4 Success index (SI)

SI, eq.(5), equally weight True positive rate (eq. 6) and specificity defined as 1 minus false positive rate (FPR), eq. (7). SI varies between 0 and 1 and its best value is 1. SI is also named modified success rate.

$$\mathrm{SI}=\frac{1}{2}\cdot\left(\frac{tp}{tp+fn}+\frac{tn}{fp+tn}\right)=\frac{1}{2}\cdot\left(\mathrm{TPR+specificity}\right) \quad (5)$$

$$\mathrm{TPR}=\frac{\mathrm{tp}}{\mathrm{tp+fn}} \quad (6) \qquad \mathrm{FPR}=\frac{\mathrm{fp}}{\mathrm{fp+tn}} \quad (7)$$

1.5 Distance to perfect classification (D2PC)

D2PC is defined in eq. (8). It measures the distance, in the plane FPR-TPR between an ideal perfect point of coordinates (0,1) and the point of the tested model (FPR,TPR). D2PC ranges in 0-1 and its best value are 0.

$$D2PC=\sqrt{\left(1-TPR\right)^2+FPR^2} \quad (8)$$

1.6 Average Index (AI)

AI, eq. (9), is the average value between four different indices: i) TPR, ii) Precision, iii) the ratio between successfully predicted stable pixels (tn) and the total number of actual stable pixels (fp+tn) and iv) the ratio between successfully predicted stable pixels (tn) and the number of simulated stable cells (fn+tn).

$$AI=\frac{1}{4}\left(\frac{tp}{tp+fn}+\frac{tp}{tp+fp}+\frac{tn}{fp+tn}+\frac{tn}{fn+tn}\right) \quad (9)$$

1.7 Heidke skill score (HSS)

The fundamental idea of a generic skill score measure is to quantify the model performance respect to set of control or reference model. Fixed a measure of model accuracy M_a, the skill score formulation is expressed in eq. (10):

$$SS=\frac{M_a - M_c}{M_{opt} - M_c} \quad (10)$$

where M_c is the control or reference model accuracy and M_{opt} is the perfect model accuracy.

SS assumes positive and negative value, if the tested model is perfect $M_a = M_{opt}$ and SS=1, if the tested model is equal to the control model than $M_a = M_c$ and SS=0.

The marginal probability of a predicted unstable pixel is (tp+fp)/n where n is the total number of pixels n=tp+fn+fp+tn. The marginal probability of a landslided unstable pixel is (tp+fn)/n.

The probability of a correct yes forecast by chance is: P1= (tp+fp) (tp+fn)/n^2. The probability of a correct no forecast by chance is: P2= (tn+fp) (tn+fn)/n^2.

In the HSS, eq. (11), the control model is a model that forecast by chance: M_c = P1+ P2, the measure of accuracy is the Accuracy (ACC) defined in eq. (12), and the M_{opt}=1.

$$HSS=\frac{2\cdot(tp\cdot tn)-(fp\cdot fn)}{(tp+fn)\cdot(fn+tn)+(tp+fp)\cdot(fp+tn)} \quad (11)$$

$$\mathrm{ACC}=\frac{\mathrm{tp+tn}}{\mathrm{tp+fn+fp+tn}} \quad (12)$$

The range of the HSS is -∞ to 1. Negative values indicate that the model provides no better results of a random model, 0 means no model skill, and a perfect model obtains a HSS of 1. HSS is also named as Cohen's kappa.

1.8 True Skill Statistic (TSS)

TSS, eq. (13), is the difference between the hit rate and the false alarm rate. It is also named Hanssen & Kuipper's Skill Score and Pierce's Skill Score. It ranges between -1 and 1 and its best value is 1. TSS equal -1 indicates that the model provides no better results of a random model. A TSS equal 0 indicates an indiscriminate model.

TSS measures the ability of the model to distinguish between landslided and non-landslided pixels. If the number of tn is large the false alarm value is relatively overwhelmed. If tn is large, as happens in landslides maps, FPR tends to zero and TSS tends to TPR. A problem of TSS is that it treats the hit rate and the false alarm rate equally, irrespective of their likely differing consequences.

$$TSS=\frac{(tp \cdot tn)-(fp \cdot fn)}{(tp+fn)\cdot(fp+tn)}=TPR-FPR \quad (13)$$

TSS is similar to Heidke, except the constraint on the reference forecasts is that they are constrained to be unbiased.

Appendix 2

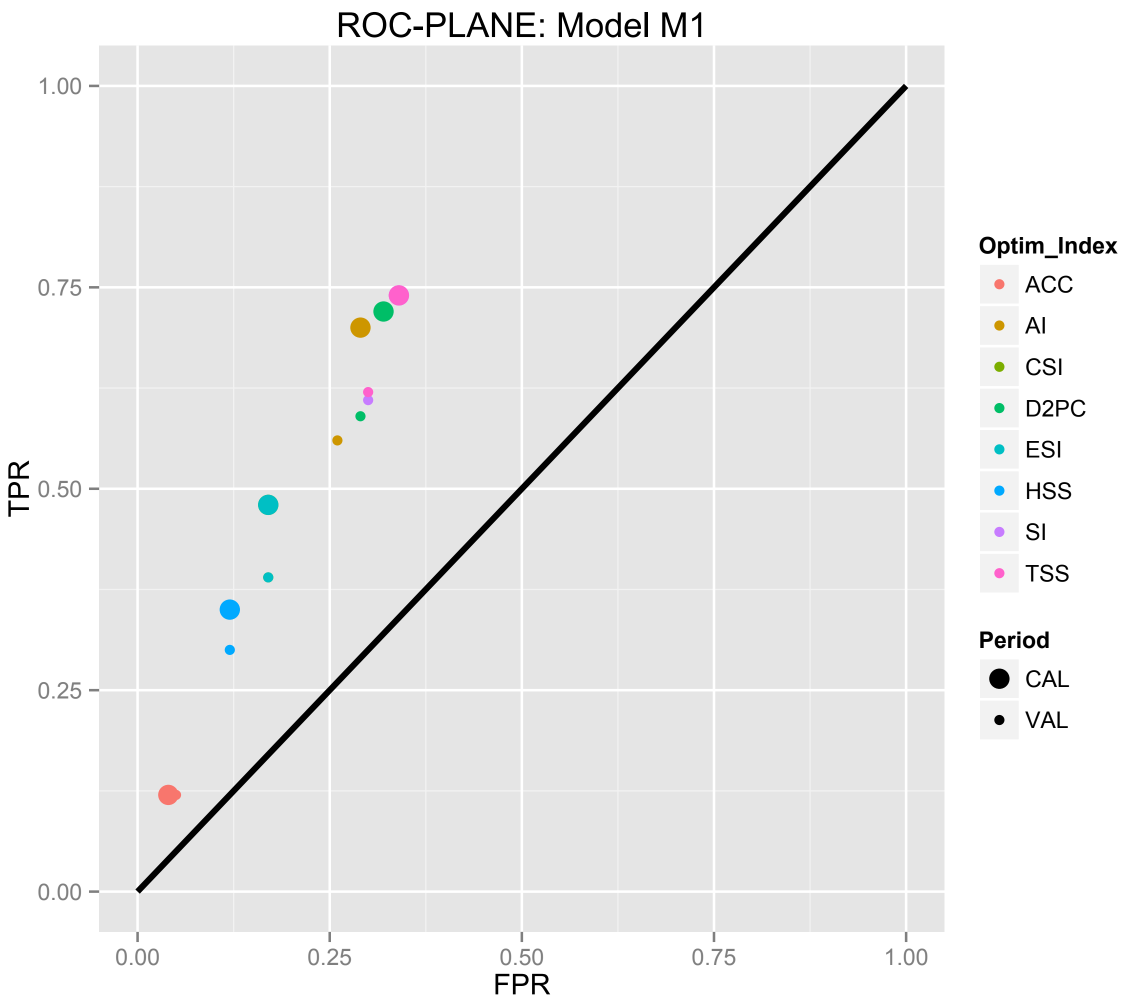


Figure A2-1: Models' performances results in the ROC plane for M1.

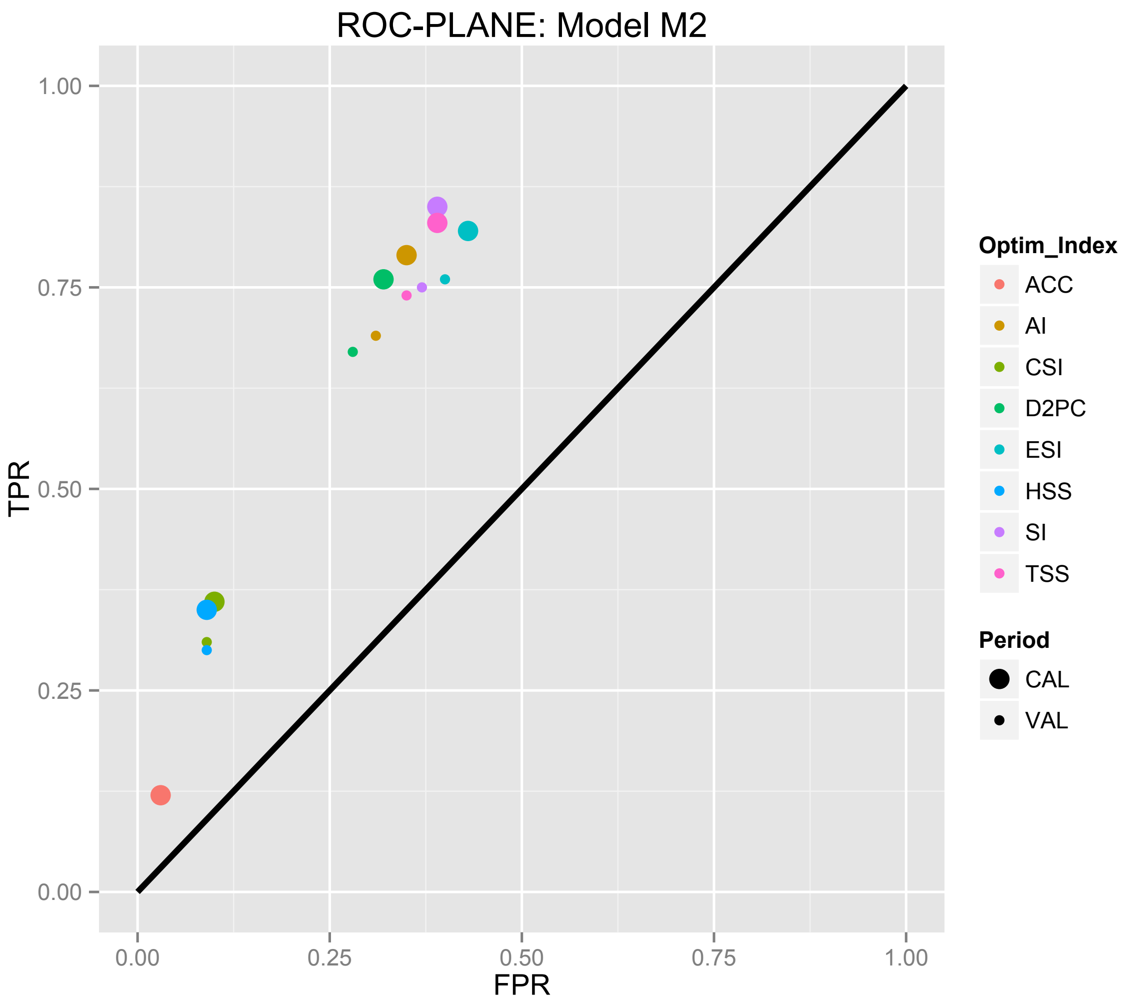


Figure A2-2: Models' performances results in the ROC plane for M2.

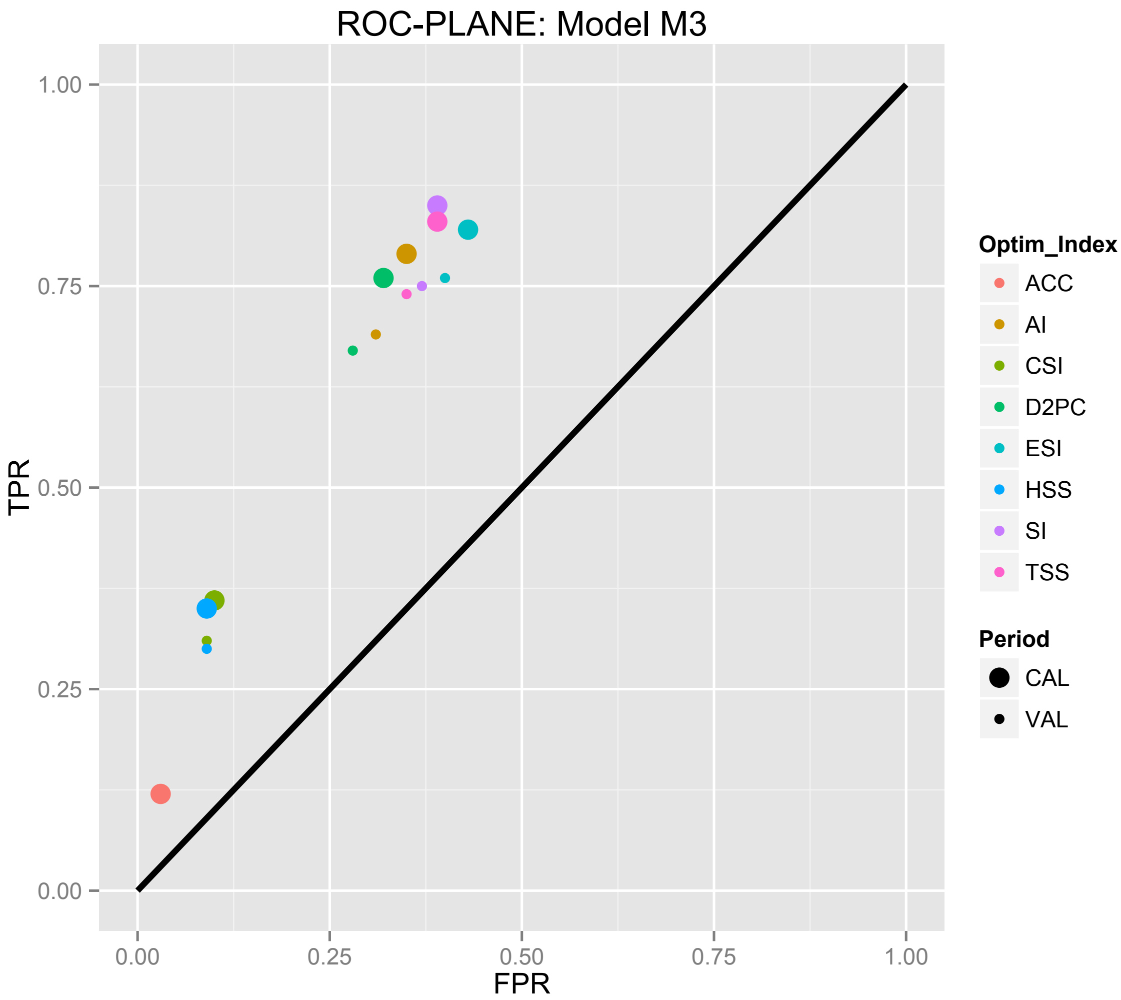


Figure A2-3: Models' performances results in the ROC plane for M3.