

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Evaluating performances of simplified physically based models for landslide susceptibility

G. Formetta, G. Capparelli, and P. Versace

University of Calabria, Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica Ponte Pietro Bucci, cubo 41/b, 87036 Rende, Italy

Received: 23 October 2015 – Accepted: 29 November 2015 – Published:

Correspondence to: G. Formetta (giuseppe.formetta@unical.it)

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

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

Rainfall induced shallow landslides cause loss of life and significant damages involving private and public properties, transportation system, etc. Prediction of shallow landslides susceptible locations is a complex task that involves many disciplines: hydrology, geotechnical science, geomorphology, and statistics. Usually to accomplish this task two main approaches are used: statistical or physically based model. Reliable models' applications involve: automatic parameters calibration, objective quantification of the quality of susceptibility maps, model sensitivity analysis. This paper presents a methodology to systemically and objectively calibrate, verify and compare different models and different models performances indicators in order to individuate and eventually select the models whose behaviors are more reliable for a certain case study.

The procedure was implemented in package of models for landslide susceptibility analysis and integrated in the NewAge-JGrass hydrological model. The package includes three simplified physically based models for landslides susceptibility analysis (M1, M2, and M3) and a component for models verifications. It computes eight goodness of fit indices by comparing pixel-by-pixel model results and measurements data. Moreover, the package integration in NewAge-JGrass allows the use of other components such as geographic information system tools to manage inputs-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 municipality. The analysis provided that among all the optimized indices and all the three models, the optimization of 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.

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



1 Introduction

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

During the last decades many methods for landslide susceptibility mapping were developed and they 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 on the basis of expert knowledge and experience, are subjective but necessary to validate quantitative methods 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 multivariate analysis, discriminant analysis, random forest to link instability factors (such as geology, soils, slope, curvature, and aspect) and past and present landslides.

Deterministic models (e.g. Montgomery and Dietrich, 1994; Lu and Godt, 2008, 2013; Borga et al., 2002; Simoni et al., 2008; Capparelli and Versace, 2011) synthesize the interaction between hydrology, geomorphology, and soil mechanics in order to physically understand and predict landslides triggering location and timing. In general, they include a hydrological and a slope stability component. The hydrological component simulates infiltration and groundwater flow processes with different degree of simplification, from steady state (e.g. Montgomery and Dietrich, 1994) to transient analysis (Simoni et al., 2008). The soil-stability component simulates the safety factor of the slope safety factor (FS) defined as ratio of stabilizing to destabilizing forces.

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Results of a landslide susceptibility analysis strongly depend on the model hypothesis, parameters values, and parameters estimation method. Problems such as the evaluation landslide susceptibility model performance, the choice of the more accurate model, and the selection of the most performing method for parameter estimation are still opened. For these reasons, a procedure that allows an objective comparisons between different models and evaluation criteria aimed to the selection of the most accurate models is needed.

Many efforts were devoted to the crucial problem of evaluating landslide susceptibility models performances (e.g. Dietrich et al., 2001; Frattini et al., 2010; Guzzetti et al., 2006). Accurate discussions about the most common quantitative measures of goodness of fit (GOF) between measured and modeled data are available in Benet et al. (2013), Jolliffe and Stephenson, (2012), Beguería (2006), Breining (2005) and references therein. We summarized them in Appendix A. Wrong classifications in landslide susceptibility analysis involve not only risk of loss of life but also economic consequences. For example locations classified as stable increase their economical value because no construction restriction will be applied, and vice-versa for locations classified as unstable.

In this work we propose an objective methodology for landslide susceptibility analysis that allows to select the most performing model based on a quantitative comparison and assessment of models prediction skills. The procedure is implemented in the open source, GIS based hydrological model, denoted as NewAge-JGrass (Formetta et al., 2014) that uses the Object Modeling System (OMS, David et al., 2013) modeling framework.

OMS is a Java based modeling framework that promotes the idea of programming by components and provides to the model developers many facilitates such as: multi-threading, implicit parallelism, models interconnection, GIS based system.

The NewAge-JGrass system, Fig. 1, contains models, automatic calibration algorithms for model parameters estimation, and methods for estimating the goodness of the models prediction. The open source GIS uDig (<http://udig.refractions.net/>) and the

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



tem. It models the whole hydrological cycle: water balance, energy balance, snow melting, etc. (Fig. 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, 2014). NewAge-JGrass is a component-based model: each hydrological process is described by a model (energy balance, evapotranspiration, run off production in Fig. 1); each model implement one or more component(s) (considering for example the model evapotranspiration in Fig. 1, the user can select between three different components: Penman–Monteith, Priestly–Taylor, and Fao); each component can be linked to the others and executed at runtime, building a model configuration. Figure 1 offers a complete picture of the system and the integration of the new LSA configuration encircled dashed red line. More precisely the LSA in the actual configuration includes two new models: a landslides susceptibility model and a model for model verification and selection. 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 steps verification procedure” (3SVP), accurately presented in Sect. 2. Moreover LSA configuration includes other two models beforehand implemented in the NewAge-JGrass system: (i) the Horton Machine for geomorphological model setup that compute input maps such as slope, total contributing area and visualize model results, and (ii) the Particle Swarm for automatic calibration. Section 2.1 presents the landslide susceptibility model and Sect. 2.2 the model selection procedure (3SVP).

2.1 Landslide susceptibility models

The landslide susceptibility models implemented in NewAge-JGrass and presented in a preliminary application in Formetta et al. (2014) are: the Montgomery and Dietrich (1994) model (M1), the Park et al. (2013) model (M3) and the Rosso et al. (2008) model (M3). The tree models derives from simplifications of the infinite slope equation (Grahm, 1984; Rosso et al., 2008; Formetta et al., 2014) for the factor of safety:

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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

where FS [-] is the factor of safety, $C = C' + C_{\text{root}}$ is the sum of C_{root} , the root strength [kNm⁻²] and C' the effective soil cohesion [kNm⁻²], φ' [-] is the internal soil friction angle, H is the soil depth [m], α [-] is the slope gradient, γ_w [kNm⁻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 solids, e [-] is the average void ratio and S_r [-] is the average degree of saturation.

The model M1 assumes 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:

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

where T [L²T⁻¹] 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 Eq. (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).

The model M2 considers both soil properties (as degree of soil saturation and void ratio) and the soil cohesion as stabilizing factors. 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 it is obtained by coupling the conservation of mass of soil water with the Darcy's law (Rosso

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc


Printer-friendly Version

Interactive Discussion



et al., 2006) providing:

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

Each component has a user interface which specifies input and output. Model input are computed in the GIS uDig integrated in the NewAge-JGrass system by using the Horton Machine package for terrain analysis (Worku  et al., 2014). Model output maps are directly imported in the GIS and available for user's visualization.

The models that we implemented present increasing degree of complexity on the theoretical assumptions for modeling landslide susceptibility. Moving from M1 to M2 soil cohesion and soil properties were considered, and moving from M2 to M3 rainfall of finite duration was used.

2.2 Automatic calibration and model verification procedure

In order to assess the models' performance we developed model that computes the most used indices for assessing the quality of a landslide susceptibility map. These are based on pixel-by-pixel comparison between 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, that 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, that correspond 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. The ROC is a methodology to assess

dex selected in Table 1 becomes an OF when it is used as 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 every single OF index for each model. We presented the results in the ROC plane in order to asses what is (are) the OF index(es) whose optimization provides best model performances. Secondly, we verified if each OF metric has own information content or if it provides information analogous to other metrics (and unessential).

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

3 Modeling framework application

The LSA presented in the paper is applied for the highway Salerno-Reggio Calabria in Calabria region (Italy), between Cosenza and Altilia. Section 3.1 describes the test-site; Sect. 3.2 describes the model parameters calibration and verification procedure; Sect. 3.3 presents the models performances correlations assessment; lastly, Sect. 3.4 presents the robustness analysis of the GOF indices used.

3.1 Site description

The test site was located in Calabria, Italy, along the Salerno-Reggio Calabria highway between Cosenza and Altilia municipalities, in the southern portion of the Crati basin (Fig. 2). The mean annual precipitation is about of 1200 mm, distributed on about 100 rainy days, and mean annual temperature of 16 °C. Rainfall peaks occur in the period October–March, during which mass wasting and severe water erosion processes are triggered (Capparelli et al., 2012; Conforti et al., 2011; Iovine et al., 2010).

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



mal parameter sets are slightly different among the models and among the optimized GOF indices for a fixed model. Moreover a compensation effect between parameter values is evident: high values of friction angles are related to low cohesion values or high values of critical rainfall are related to high values of soil resistance parameters. Considering the model M1, transmissivity value ($74 \text{ m}^2 \text{ day}^{-1}$) optimizing ACC is much lower compared to the transmissivity values obtained optimizing the other index (around $140 \text{ m}^2 \text{ day}^{-1}$). Similar behavior is observed for the optimal rainfall value which is $148 [\text{mm day}^{-1}]$ optimizing ACC and around $70 [\text{mm day}^{-1}]$ optimizing the other indices. Considering the model M2, the optimal transmissivity and rainfall values optimizing CSI ($10 [\text{m}^2 \text{ day}^{-1}]$ and $95 [\text{mm day}^{-1}]$), are much lower compared the values obtained optimizing the other indices (around $50 [\text{m}^2 \text{ day}^{-1}]$ and $250 [\text{mm day}^{-1}]$ in average). For the model M3, instead, optimal parameters present the same order of magnitude for all optimized indices. This suggests that the variability of the optimal parameter values for model 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 model output and actual landslides for both calibration and verification dataset. Results were presented in Table 4, for all three models M1, M2 and M3. Those points were reported in the ROC plane in order to visualize in a unique graph the effects of the optimised objective function on model performances. This procedure was repeated for the three models. ROC planes considering all the GOF indices and all three models are included in Figs. B1–B3 both for calibration and for verification period. For model M2 and M3 is clear that ACC, HSS, and CSI provides the less performing models results. This is true also for model M1, even if, differently from M2 and M3, there is not a so clear separation between the performances provided by ACC, HSS, and CSI and the remaining indices.

Among the results provided in Table 4, we focused our attention only on the GOF indices whose optimization satisfies the condition: $\text{FPR} < 0.4$ and $\text{TPR} > 0.7$. This choice

MP_j vectors. The ellipse's eccentricity is scaled according to the correlation value: the more is prominent as the less the vector are correlated; if ellipse leans towards the right correlation is positive and if it leans to the left, it is negative.

All indices present a positive correlation among each other independent of the model used. Moreover strong correlations between the MP vectors of AI, D2PC, SI and TSS are evident in Fig. 4. This confirms that an optimization of AI, D2PC, SI and TSS provide quite similar model performances, and this is independent of the model used. On the other hand the remaining GOF indices give quite different information from the previous four indices, but they gave worse performances in first step analysis. Thus in the case study using one of the four best GOF can be enough for parameter estimation.

3.4 Models sensitivity assessment

In this step we focused the attention on the models M2 and M3 and we performed a parameter sensitivity analysis. Let's assume to consider model M2 and the optimal parameter set computed by optimizing the Critical Success Index (CSI). Moreover let's assume to consider the cohesion model parameter, the procedure evolves according 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 picked up from a uniform distribution with lower and upper bound defined in Table 1. With this procedure 1000 model parameter sets were defined and used to execute the model.

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- 1000 values of the selected GOF index (CSI), computed by comparing model outputs with 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 where presented in Figs. 5 and 6 for model M2 and M3 respectively.

Each column of the figures represents one optimized index and has a number of boxplot equal to the number of model's parameters (5 for M2 and 6 for M3). Each boxplot represents the range of variation of the optimized index due a certain model parameters change. The more narrow are the boxplot for a given optimized index the less sensitive is the model to that parameter. For both M2 and M3 the parameter set obtained by optimizing AI and SI shows the less sensitive behavior for almost all parameters. In this case a model parameter perturbation does not influence much the model performances. On the contrary, the models whit parameters obtained by optimizing ACC, TSS, and D2PC are the more sensitive to the parameters variations and this is reflected in much more evident changing of model performances.

3.5 Models selections and susceptibility maps

The selection of the more appropriate model for computing landslide susceptibility maps is based on what we learn forms the previous steps. In the first step we learn that (i) optimization of AI, D2PC, SI and TSS outperform the remaining indices and (ii) models M2 and M3 provides more accurate results compared to M1. The second step suggests that overall models results obtained by optimizing AI, D2PC, SI and TSS are similar each others. Lastly, the third step show that models performance derived from the optimization of AI and SI are the less sensible to input variations compared to D2PC and TSS. This behavior could be due the formulation of AI and SI that gives much more weight to the true positive compared to D2PC and TSS.

In particular for our application, the model M3 whit parameters obtained by optimizing D2PC was the most sensitive to the parameter variation avoiding an “insensitive” or

Evaluating performances of simplified physically based models

G. Formetta et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



$$SI = \frac{1}{2} \times \left(\frac{tp}{tp + fn} + \frac{tn}{fp + tn} \right) = \frac{1}{2} \times (TPR + \text{specificity}) \quad (A4)$$

$$TPR = \frac{tp}{tp + fn} \quad (A5)$$

$$FPR = \frac{fp}{fp + tn} \quad (A6)$$

A4 Distance to perfect classification (D2PC)

5 D2PC is defined in Eq. (A7). 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.

$$D2PC = \sqrt{(1 - TPR)^2 + FPR^2} \quad (A7)$$

A5 Average Index (AI)

10 AI, Eq. (A8), 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).

$$AI = \frac{1}{4} \left(\frac{tp}{tp + fn} + \frac{tp}{tp + fp} + \frac{tn}{fp + tn} + \frac{tn}{fn + tn} \right) \quad (A8)$$

15 A6 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

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



M_a , the skill score formulation is expressed in Eq. (A9):

$$SS = \frac{M_a - M_c}{M_{opt} - M_c} \quad (A9)$$

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. (A10), 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. (A11), and the $M_{opt} = 1$.

$$HSS = \frac{2 \times (tp \times tn) - (fp \times fn)}{(tp + fn) \times (fn + tn) + (tp + fp) \times (fp + tn)} \quad (A10)$$

$$ACC = \frac{tp + tn}{tp + fn + fp + tn} \quad (A11)$$

The range of the HSS is $-\infty$ to 1. Negative values indicate that indicates 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.

A7 True Skill Statistic (TSS)

TSS, Eq. (A12), is the difference between the hit rate and the false alarm rate. It is also named Hanssen and Kuipper's Skill Score and Pierce's Skill Score. It ranges between

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



-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-landsided 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 tests the hit rate and the false alarm rate equally, irrespective of their likely differing consequences.

$$TSS = \frac{(tp \times tn) - (fp \times fn)}{(tp + fn) \times (fp + tn)} = TPR - FPR \quad (A12)$$

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

Acknowledgements. This research was funded by PON Project No. 01_01503 “Integrated Systems for Hydrogeological Risk Monitoring, Early Warning and Mitigation Along the Main Lifelines”, CUP B31H11000370005, in the framework of the National Operational Program for “Research and Competitiveness” 2007–2013.

References

Abera, W., Bonello, A., Franceschi, S., Formetta, G., and Rigon, R.: “The uDig Spatial Toolbox for hydro-geomorphic analysis”, in: *Geomorphological Techniques*, v. 4, n. 1, 1–19, available at: http://www.geomorphology.org.uk/sites/default/files/geom_tech_chapters/2.4.1_GISToolbox.pdf (last access: December 2015), 2014.

Boggs, R., Savage, W., and Godt, J.: TRIGRS, a Fortran Program for Transient Rainfall Infiltration and Grid-Based Regional Slope-Stability Analysis, US Geological Survey Open Report 424, US Geological Survey, Golden, CO, 61 pp., 2002.

Beguiría, S.: Validation and evaluation of predictive models in hazard assessment and risk management, *Nat. Hazards*, 37, 315–329, 2006.

Bennett, N. D., Croke, B. F., Guariso, G., Guillaume, J. H., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T., Norton, J. P., Perrin, C., and Pierce, S. A.: Characterising performance of environmental models, *Environ. Modell. Softw.*, 40, 1–20, 2013.

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Borga, M., Dalla Fontana, G., and Cazorzi, F.: Analysis of topographic and climatic control on rainfall-triggered shallow landsliding using a quasi-dynamic wetness index, *J. Hydrol.*, 268, 56–71, 2002.

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

Brown, C. D. and Davis, H. T.: Receiver operating characteristics curves and related decision measures: a tutorial, *Chemometr. Intell. Lab.*, 80, 24–38, 2006.

Capparelli, G. and Versace, P.: FLalR and SUSHI: two mathematical models for early warning of landslides induced by rainfall, *Landslides*, 8, 67–79, 2011.

Capparelli, G., Iaquinata, P., Iovine, G. G. R., Terranova, O. G., and Versace, P.: Modelling the rainfall-induced mobilization of a large slope movement in northern Calabria, *Nat. Hazards*, 61, 247–256, 2012.

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

Catani, F., Casagli, N., Ermini, L., Righini, G., and Menduni, G.: Landslide hazard and risk mapping at catchment scale in the Arno River basin, *Landslides*, 2, 329–342, 2005.

Chung, C.-J. F., Fabbri, A. G. and van Westen, C. J.: Multivariate regression analysis for landslide hazard zonation, in: *Geographical Information Systems in Assessing Natural Hazards*, edited by: Carrara, A. and Guzzetti, F., Kluwer Academic Publishers, Dordrecht, 107–34, 1995.

Colella, A., De Boer, P. L., and Nio, S. D.: Sedimentology of a marine intermontane Pleistocene Gilbert-type fan-delta complex in the Crati Basin, Calabria, southern Italy, *Sedimentology*, 34, 721–736, 1987.

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

Conforti, M., Pascale, S., Robustelli, G., and Sdao, F.: 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, 2014.

Evaluating performances of simplified physically based models

G. Formetta et al.

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

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

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Evaluating performances of simplified physically based models

G. Formetta et al.

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 and verification dataset. In the rows for which the condition $FPR < 0.4$ and $TPR > 0.7$ is verified.

Period	Optim. index	MODEL: M1		MODEL: M2		MODEL: M3	
		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

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table A1. 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
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

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Model setup
Interpolation tools
Energy balance
Water balance
Calibration

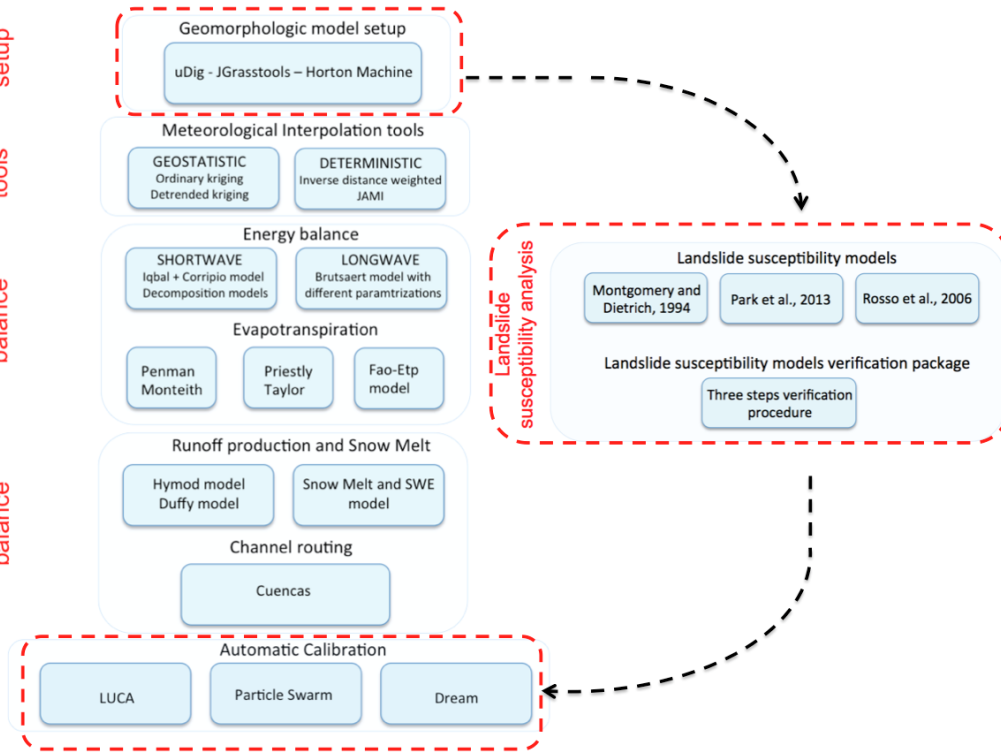


Figure 1. Integration of the Landslide susceptibility analysis system in NweAge-JGrass hydrological model.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



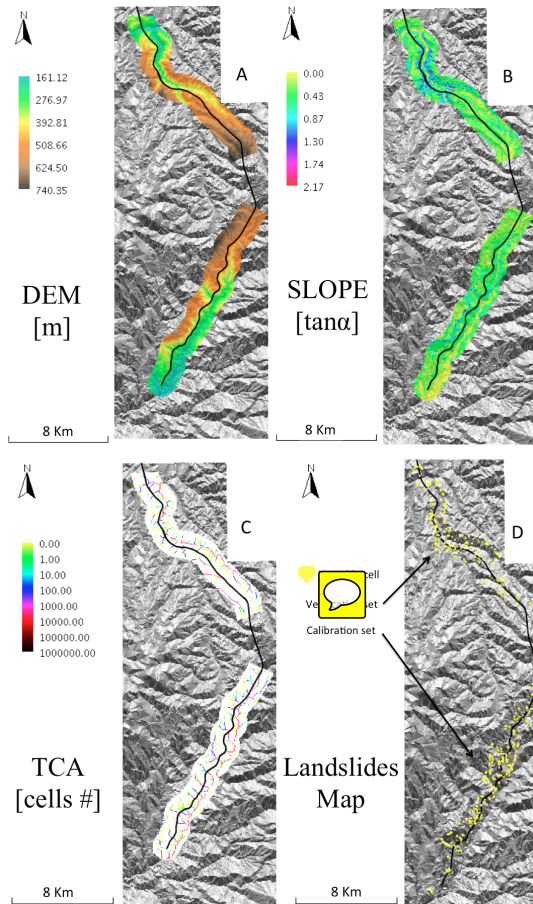


Figure 2. Test site. **(a)** Digital elevation model (DEM) [m], **(b)** slope [-] expressed as tangent of the angle, **(c)** total contributing area (TCA) expressed as number of draining cells and **(d)** map of actual landslides.

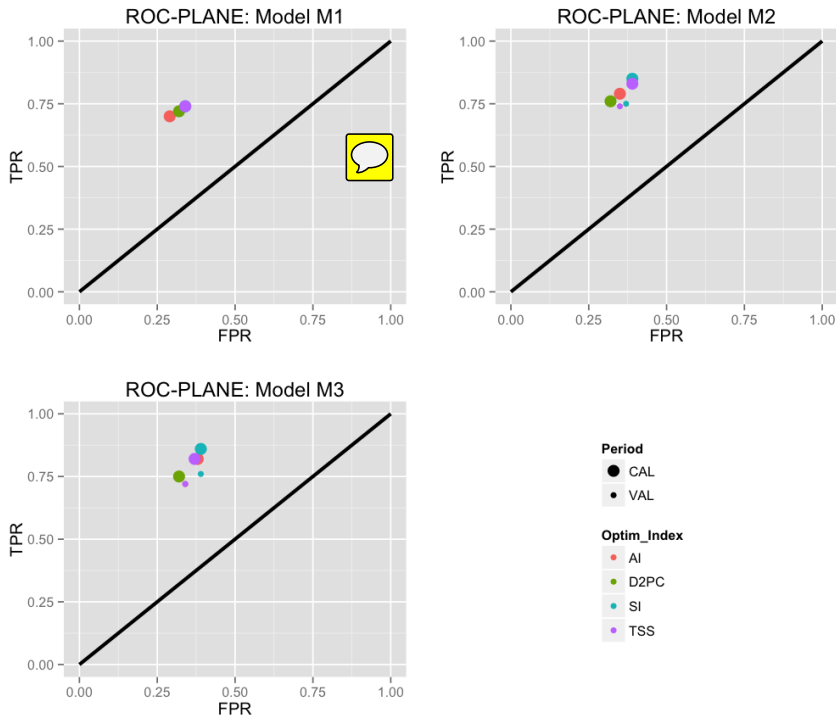


Figure 3. Models' performances results in the ROC plane for M1, M2 and M3. Only GOF indices whose optimization provides $FPR < 0.4$ and $TPR > 0.7$ were reported.

Evaluating performances of simplified physically based models

G. Formetta et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



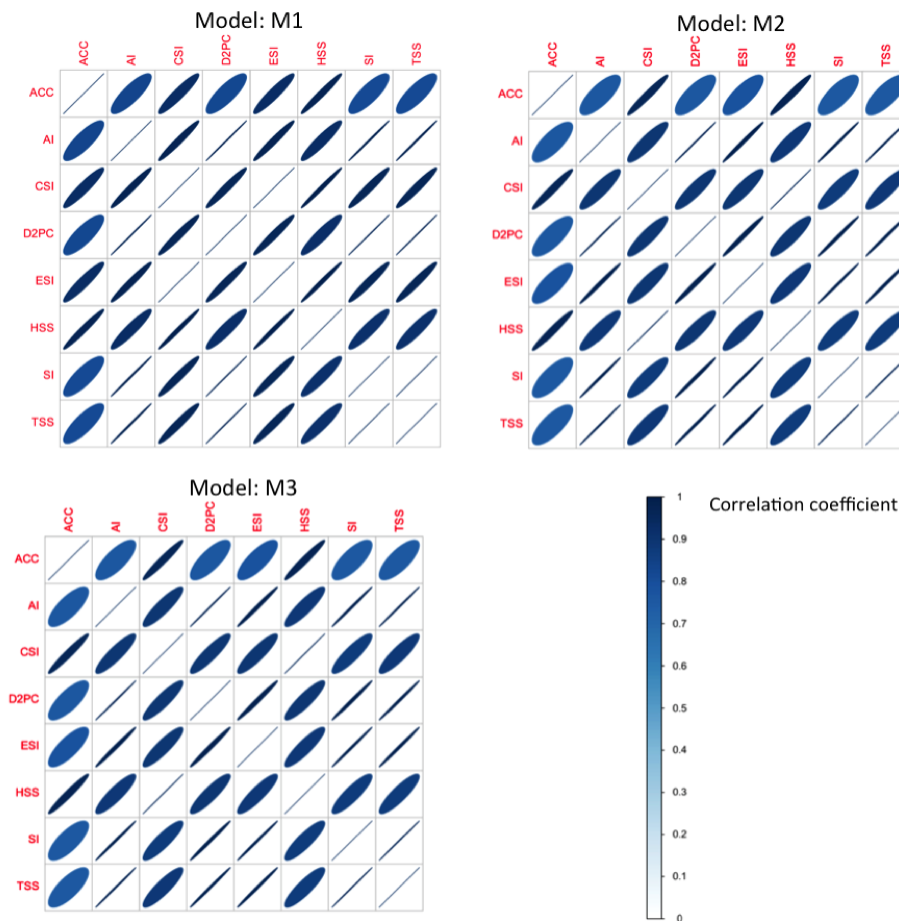


Figure 4. Correlation plot between models' performance (MP) vector computed by optimizing all GOF indices in turn. Results are reported for each model: M1, M2 and M3.

Evaluating performances of simplified physically based models

G. Formetta et al.

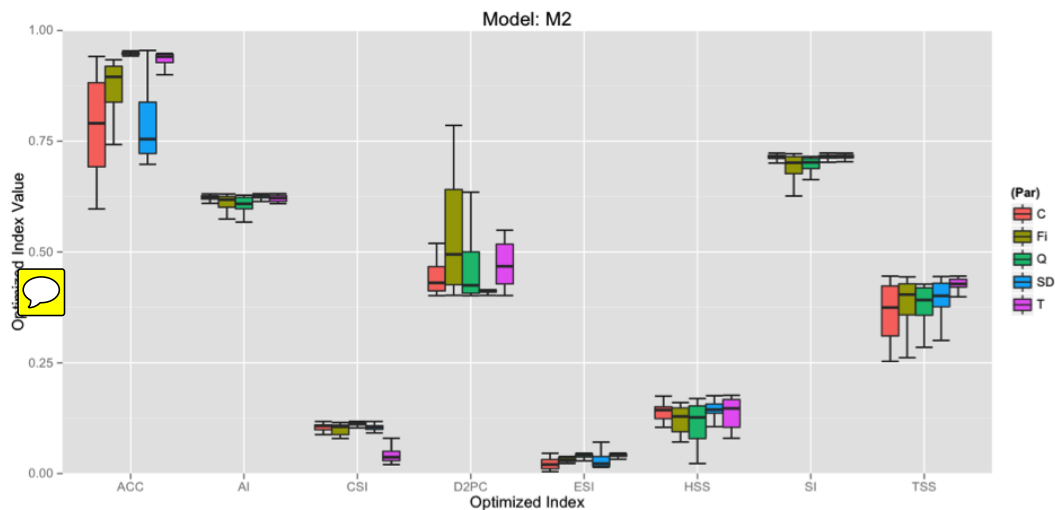


Figure 5. Model M2 parameters sensitivity analysis.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

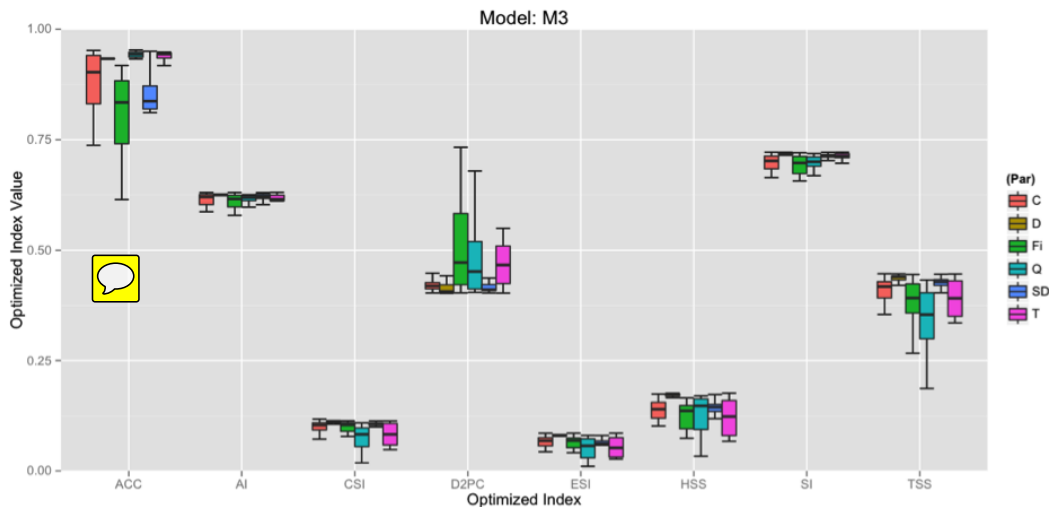


Figure 6. Model M3 parameters sensitivity analysis.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



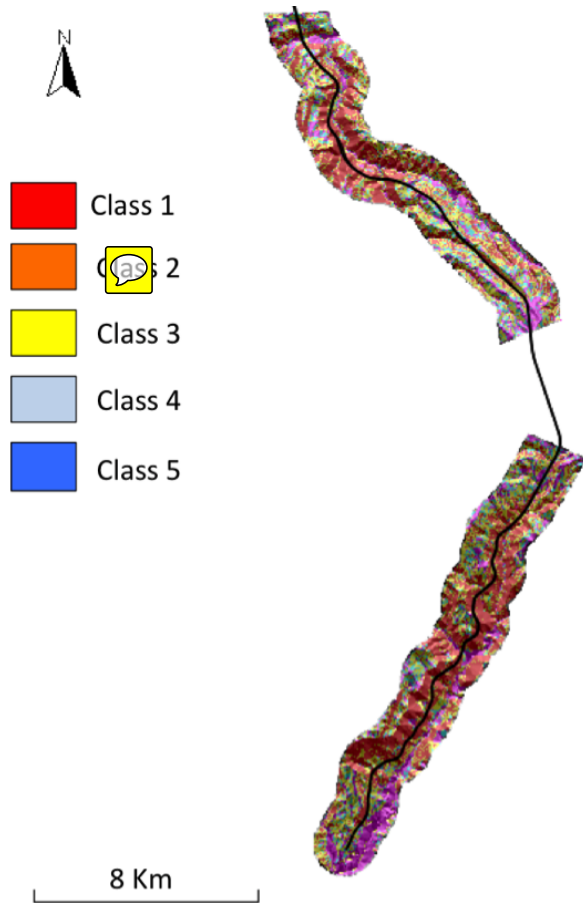


Figure 7. Landslide susceptibility maps using model M3 and parameter set obtained by optimising D2PC.

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



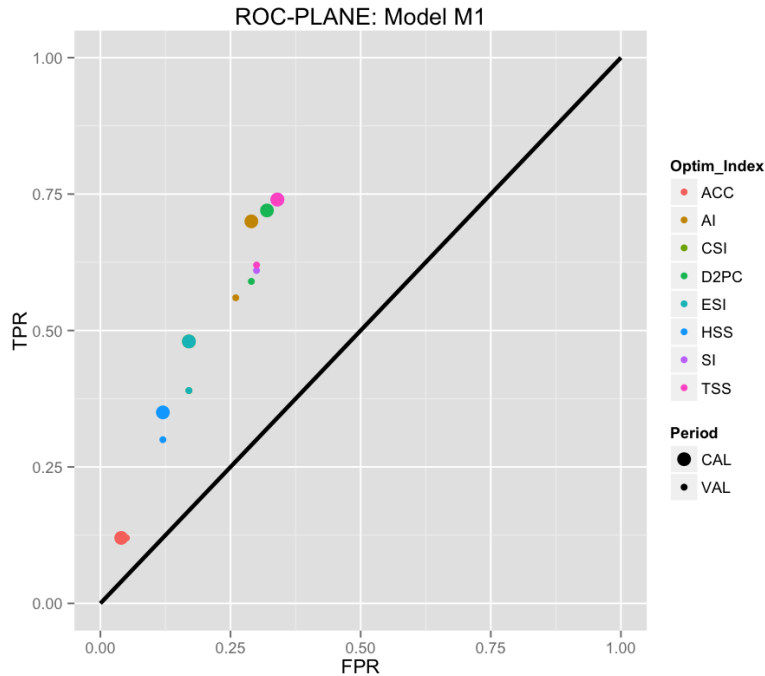


Figure B1. Models' performances results in the ROC plane for M1.

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

[Title Page](#)

[Abstract](#) | [Introduction](#)

[Conclusions](#) | [References](#)

[Tables](#) | [Figures](#)

[⏪](#) | [⏩](#)

[⏴](#) | [⏵](#)

[Back](#) | [Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



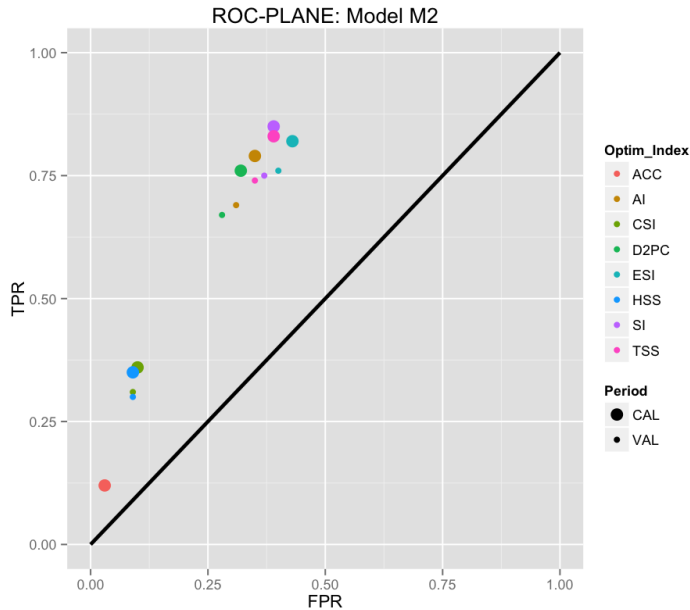


Figure B2. Models' performances results in the ROC plane for M2.

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

[Title Page](#)

[Abstract](#) | [Introduction](#)

[Conclusions](#) | [References](#)

[Tables](#) | [Figures](#)

[◀](#) | [▶](#)

[◀](#) | [▶](#)

[Back](#) | [Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



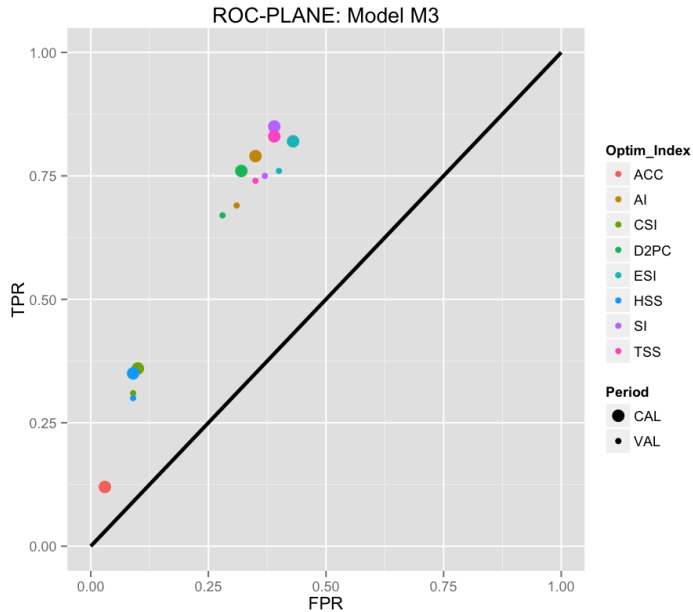


Figure B3. Models' performances results in the ROC plane for M3.

HESSD

19, 1–40, 2015

Evaluating performances of simplified physically based models

G. Formetta et al.

[Title Page](#)

[Abstract](#) | [Introduction](#)

[Conclusions](#) | [References](#)

[Tables](#) | [Figures](#)

[⏪](#) | [⏩](#)

[⏴](#) | [⏵](#)

[Back](#) | [Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

