

1 Evaluating Performances of Simplified Physically Based
2 Models for Landslide Susceptibility.

3
4 **Giuseppe Formetta, Giovanna Capparelli and Pasquale Versace**

5
6 *University of Calabria Dipartimento di Ingegneria Informatica, Modellistica,*
7 *Elettronica e Sistemistica Ponte Pietro Bucci, cubo 41/b, 87036 Rende, Italy*

8 *(giuseppe.formetta@unical.it, giovanna.capparelli@unical.it,*
9 *pasquale.versace@unical.it)*

10
11 **Abstract:** Rainfall induced shallow landslides can lead to loss of life and significant
12 damage to private and public properties, and transportation systems, etc. Predicting
13 locations that might be susceptible to shallow landslides is a complex task and
14 involves many disciplines: hydrology, geotechnical science, geology, hydrogeology,
15 geomorphology, and statistics. Two main approaches are commonly used: statistical
16 or physically based models. Reliable model applications involve automatic parameter
17 calibration, objective quantification of the quality of susceptibility maps, and model
18 sensitivity analyses. This paper presents a methodology to systemically and
19 objectively calibrate, verify and compare different models and model performance
20 indicators in order to identify and select the models whose behaviors are the most
21 reliable for particular case studies.

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

30 The system was applied for a case study in Calabria (Italy) along the Salerno-Reggio
31 Calabria highway, between Cosenza and Altilia. The area is extensively subject to
32 rainfall-induced shallow landslides mainly because of its complex geology and

33 climatology. The analysis was carried out considering all the combinations of the
34 eight optimized indices and the three models. Parameter calibration, verification, and
35 model performance assessment were performed by a comparison with a detailed
36 landslide inventory map for the area. The results showed that the index distance to
37 perfect classification in the receiver operating characteristic plane (D2PC) coupled
38 with model M3 is the best modeling solution for our test case.

39

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

41

42 **1 INTRODUCTION**

43

44 Landslides are one of the main dangerous geo-hazards worldwide and constitute a
45 serious menace for public safety leading to human and economic losses (Park
46 2011). Geo-environmental factors such as geology, land-use, vegetation, climate,
47 and increasing populations may increase the occurrence of landslides (Sidle and
48 Ochiai 2006). Landslide susceptibility assessments, i.e. the likelihood of a landslide
49 occurring in an area on the basis of local terrain conditions (Brabb, 1984), is not only
50 crucial for an accurate landslide hazard quantification but also a fundamental tool for
51 the environmental preservation and responsible urban planning (Cascini et al.,
52 2005).

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

56 Qualitative methods, based on field campaigns and expert knowledge and
57 experience, are subjective but necessary to validate quantitative method results.

58 Quantitative methods include statistical and physically based methods. Statistical
59 methods (e.g. Naranjo et al., 1994; Chung et al. 1995; Guzzetti et al., 1999; Catani
60 et al., 2005) use different approaches such as bivariate statistics, multivariate
61 analysis, discriminant analysis, random forest to link instability factors (such as
62 geology, soil, slope, curvature, and aspect) with past and present landslides.
63 Bivariate statistical methods ignore the interdependence of instability factors
64 whereas multivariate analysis is able to statistically consider their interactions. Other
65 data-driven methods for landslide susceptibility analysis include the use of neural

66 networks (Pradhan, 2011; Conforti et al., 2014), support vector machines (Pradhan,
67 2013 and citations therein), and Bayesian networks (Lee et al., 2002). Deterministic
68 models (e.g. Montgomery and Dietrich, 1994; Lu and Godt, 2008; Borga et al., 2002;
69 Simoni et al., 2008; Capparelli and Versace, 2011; Lu and Godt, 2013) synthesize
70 the interaction between hydrology, geomorphology, and soil mechanics in order to
71 physically understand and predict the location and timing that trigger landslides.
72 These models generally include a hydrological and a slope stability component. The
73 hydrological component simulates infiltration and groundwater flow processes with
74 different degrees of simplification, from steady state (e.g. Montgomery and Dietrich,
75 1994) to transient analyses (Simoni et al., 2008). The soil-stability component
76 simulates the slope safety factor (FS) defined as the ratio of stabilizing to
77 destabilizing forces. One of the main advantages of data-driven methods for
78 landslide susceptibility is that they can be easily applied in wide areas while
79 deterministic models are in general applied in local analyses. The latter are more
80 computationally expensive and require detailed input data and parameters, which
81 often involve high uncertainty. On the other hand, data-driven methods assume that
82 landslides are caused by the same combination of instability factors overall the study
83 area, whereas deterministic models enable different triggering mechanisms to be
84 understood and investigated.

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

92 Much effort has been devoted to the crucial problem of evaluating landslide
93 susceptibility model performances (e.g Dietrich et al., 2001; Frattini et al., 2010 and
94 Guzzetti et al., 2006). Accurate discussions about the most common quantitative
95 measures of goodness of fit (GOF) between measured and modeled data are
96 discussed in Bennet et al., (2013), Jolliffe and Stephenson, (2012), Beguería (2006),
97 Brenning (2005) and references therein. We have summarized them in Appendix 1.
98 Usually one of these indices is selected and used as an objective function (OF) in

99 combination with a calibration algorithm in order to obtain the optimal set of model
100 parameters. However, in most cases the selection of the OF is not justified or
101 compared with other options.

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

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

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

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

127 The NewAge-JGrass system, Fig. 1, contains models, automatic calibration
128 algorithms for model parameter estimation, and methods for estimating the
129 goodness of the models prediction. The open source GIS uDig
130 (<http://udig.refractions.net/>) and the uDig-Spatial Toolbox (Abera et al., (2014),
131 <https://code.google.com/p/jgrasstools/wiki/JGrassTools4udig>) are used as a

132 visualization and input/out data management system. The OMS framework has been
133 previously used as the core for landslides modeling (Formetta et al., 2016; Formetta
134 et al., 2015). These studies deal with real time early warning systems for landslide
135 risks and involve 3D physically based hydrological modeling of very small
136 catchments (up to around 20 km²). In contrast, the current application focuses on
137 wider areas landslide susceptibility assessments using completely different
138 physically based models which are presented in the next section.

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

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

156

157 **2 MATERIALS AND METHODS**

158

159 **2.1 Modelling Framework**

160

161 The landslide susceptibility analysis (LSA) is implemented in the context of NewAge-
162 JGrass (Formetta et al., 2014), an open source large-scale hydrological modeling
163 system. It models the whole hydrological cycle: water balance, energy balance, snow
164 melting, etc. (Figure 1). The system implements hydrological models, automatic

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

185

186 **2.2 Landslide susceptibility models**

187

188 The landslide susceptibility models implemented in NewAge-JGrass and presented
 189 in a preliminary application in Formetta et al., 2015 consist of the Montgomery and
 190 Dietrich (1994) model (M1), the Park et al. (2013) model (M2) and the Rosso et al.
 191 (2006) model (M3). The three models derive from simplifications of the infinite slope
 192 equation (Grahm J., 1984, Rosso et al., 2006, Formetta et al., 2014) for the factor of
 193 safety:

194

$$195 \quad FS = \frac{C \cdot (1+e)}{[G_s + e \cdot S_r + w \cdot e \cdot (1-S_r)] \cdot \gamma_w \cdot H \cdot \sin \alpha \cdot \cos \alpha} + \frac{[G_s + e \cdot S_r - w \cdot (1+e \cdot S_r)]}{[G_s + e \cdot S_r + w \cdot e \cdot (1-S_r)]} \cdot \frac{\tan \varphi'}{\tan \alpha} \quad (1)$$

196

197 where FS [-] is the factor of safety, $C=C'+C_{root}$ is the sum of C_{root} , the root strength
 198 [kN/m²] and C' the effective soil cohesion [kN/m²], φ' [-] is the internal soil friction
 199 angle, H is the soil depth [m], α [-] is the slope angle, γ_w [kN/m³] is the specific
 200 weight of water, and $w=h/H$ [-] where h [m] is the water table height above the failure
 201 surface [m], G_s [-] is the specific gravity of soil, e [-] is the average void ratio and S_r
 202 [-] is the average degree of saturation.

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

206

$$207 \quad w = \frac{h}{H} = \min\left(\frac{Q}{T} \cdot \frac{TCA}{b \cdot \sin \alpha}, 1.0\right) \quad (2)$$

208

209 where T [L²/T] is the soil transmissivity defined as the product of the soil depth and
 210 the saturated hydraulic conductivity, b [L] is the length of the contour line.
 211 Substituting eq. (2) in (1) the model is solved for Q/T assuming $FS=1$ and stable and
 212 unstable sites are defined using threshold values on $\log(Q/T)$ (Montgomery and
 213 Dietrich, 1994).

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

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

222

$$223 \quad w = \begin{cases} \frac{Q}{T} \cdot \frac{TCA}{b \cdot \sin \alpha} \cdot \left[1 - \exp\left(-\frac{e+1}{e \cdot (1-S_r)} \cdot \frac{t}{T} \cdot \frac{TCA}{b \cdot \sin \alpha} \cdot H\right) \right] & \text{if } \frac{t}{T} \cdot \frac{TCA}{b \cdot \sin \alpha} \cdot H \leq -\frac{e \cdot (1-S_r)}{1+e} \cdot \ln\left(1 - \frac{T \cdot b \cdot \sin \alpha}{TCA \cdot Q}\right) \\ 1 & \text{if } \frac{t}{T} \cdot \frac{TCA}{b \cdot \sin \alpha} \cdot H > -\frac{e \cdot (1-S_r)}{1+e} \cdot \ln\left(1 - \frac{T \cdot b \cdot \sin \alpha}{TCA \cdot Q}\right) \end{cases} \quad (3)$$

224

225 These models are suitable for shallow translational landslides controlled by
226 groundwater flow convergence. Shallow landslides usually have a very low ratio
227 between the maximum depth (D) and the length (L) of scar ($D/L < 0.1$, Casadei et al.,
228 2003), involve a small volume of the colluvial soil mantle and present a generally
229 translational failure mechanism (Milledge et al., 2014).

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

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

239

240 **2.3 Automatic calibration and model verification procedure**

241

242 In order to assess the models' performance we developed a model that computes
243 the most common indices for assessing the quality of a landslide susceptibility map.

244 These indices are based on a pixel-by-pixel comparison between the observed
245 landslide map (OL) and predicted landslides (PL). They are binary maps with
246 positive pixels corresponding to "unstable" ones, and negative pixels that correspond
247 to "stable" ones. Therefore, four types of outcomes are possible for each cell. A pixel
248 is a true-positive (tp) if it is mapped as "unstable" both in OL and in PL, which is a
249 correct alarm with well predicted landslide. A pixel is a true-negative (tn) if it is
250 mapped as "stable" both in OL in PL, which corresponds to a well predicted stable
251 area. A pixel is a false-positive (fp) if it is mapped as "unstable" in PL, but is "stable"
252 in OL; that is a false alarm. A pixel is a false-negative (fn) if it is mapped as "stable"
253 in PL, but is "unstable" in OL, that is a missed alarm. The concept of the Receiver
254 Operator Characteristic (ROC, Goodenough et al., 1974) graph is based on the
255 values assumed by tp, fp, tn. ROCs are used to assess the performance of models
256 which provides results assigned to one of two classes. The ROC graph is widely
257 used in many scientific fields such as medicine (Goodenough et al., 1974),

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

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

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

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

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

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

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

298

299 **2.4 Site Description**

300

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

308 In the study area the topographic elevation has an average value of around 450 m
309 a.s.l., with a maximum value of 730 m a.s.l. Slopes, computed from the 10 meters
310 resolution digital elevation model, range from 0° to 55°, while the average is about
311 26°.

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

321 In the study area the second unit outcrops. A topsoil of about 1.5 - 2.0 m lies on
322 sandy-gravelly and sandy deposits, which are generally well-stratified. Soils range
323 from Alfisols (i.e. highly mature soils) to Inceptisols and Entisols (i.e. poorly

324 developed soils). Due to the combination of such climatic, geo-structural, and
325 geomorphological features the test site is one of the most landslide prone areas in
326 Calabria (Conforti et al., 2014; Carrara and Merenda, 1976; Iovine et al., 2006,).
327 Mass movements were analyzed from 2006 to 2013 by integrating aerial
328 photography interpretation acquired in 2006, 1:5000 scale topographic maps
329 analysis, and an extensive field survey.
330 All the data were digitized and stored in a GIS database (Conforti et al., 2014) and
331 the result was the map of occurred landslides, presented in Figure 2,D. Digital
332 elevation model, slope and total contributing area (TCA) maps are presented in
333 Figures 2, A, B, and C respectively. In order to perform model calibration and
334 verification, the dataset of occurred landslides was divided in two parts one used for
335 calibration (located at bottom of Figure 2,D) and one for validation (located in the
336 upper part of Figure 2,D). The landslide inventory map refers only to the initiation
337 area of the landslides. This leads to a fair comparison with the landslide models that
338 provide only the triggering point and does not include a runout model for landslides
339 propagation.

340

341 **3 RESULTS AND DISCUSSION**

342

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

349

350 **3.1 Model calibration and verification**

351

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

357 The component PSO provides eighth best parameter sets, one for each optimized
358 GOF indices. Values for each model (M1, M2 and M3) are presented in Table 3.
359 Optimal parameter sets differ slightly among the models and among the optimized
360 GOF indices for a given model. In addition a compensation effect between the
361 parameter values is evident. High values of friction angle are related to low cohesion
362 values; high values of critical rainfall are related to high values of soil resistance
363 parameters. For the model M1, the transmissivity value (74 m²/d) optimizing ACC is
364 much lower than the transmissivity values obtained by optimizing the other indices
365 (around 140 m²/d). Similar behavior was observed for the optimal rainfall value
366 which is 148 [mm/d] optimizing ACC, and around 70 [mm/d] optimizing the other
367 indices. For the model M2, the optimal transmissivity and rainfall values optimizing
368 CSI (10 [m²/d] and 95 [mm/d]), are much lower than the values obtained by
369 optimizing the other indices (around 50 [m²/d] and 250 [mm/d] in average). For the
370 model M3, on the other hand, optimal parameters present the same order of
371 magnitude for all the optimized indices. This suggests that the variability of the
372 optimal parameter values for models M1 and M2 could be due to compensate the
373 effects of important physical processes neglected by those models.

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

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

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

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

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

407

408 **3.2 Correlations assessment of the models performances**

409

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

413 Executing a model using one of the eight parameters set (assuming, for example,
 414 the one obtained by optimizing CSI) enables all the remaining GOF indices to be
 415 computed, which we indicate as CSI_{CSI} , ACC_{CSI} , HSS_{CSI} , TSS_{CSI} , AI_{CSI} , SI_{CSI} ,
 416 $D2PC_{CSI}$, ESI_{CSI} , both for calibration and for verification dataset. Let us denote this
 417 vector with the name ***MP***_{CSI}: the model performance (***MP***) vector computed using the
 418 parameter set that optimizes CSI. ***MP***_{CSI} has 16 elements, 8 for the calibration and 8
 419 for the validation dataset. Repeating the same procedure for all eight GOF indices it
 420 gives: ***MP***_{ACC}, ***MP***_{ESI}, ***MP***_{SI}, ***MP***_{D2PC}, ***MP***_{TSS}, ***MP***_{AI}, ***MP***_{HS}. Figure 4 presents the
 421 correlation plots (Murdoch and Chow, 1996) between all ***MP*** vectors, for each model

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

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

436

437 **3.3 Models sensitivity assessment**

438

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

- 443 • The starting parameter values are the optimal values derived from the
444 optimization of the CSI index;
- 445 • All the parameters except the analyzed parameter (cohesion) were kept
446 constant and equal to the optimal parameter set;
- 447 • 1000 random values of the analyzed parameter (cohesion) were selected
448 from a uniform distribution with the lower and upper bound defined in Table 1.
449 With this procedure 1000 model parameter sets were defined and used to
450 execute the model.
- 451 • 1000 values of the selected GOF index (CSI), computed by comparing model
452 outputs with the measured data, were used to compute a boxplot of the
453 parameter C and optimized index CSI.

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

472

473 **3.4 Models selections and susceptibility maps**

474

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

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

487 optimal parameter set will in fact accommodate any parameters, input data, or
488 measured data variations responding to these changes with a variation in model
489 performance.

490 We thus used the combination of model M3 with parameters obtained by optimizing
491 D2PC in order to compute the final susceptibility maps in Figure 7. Categories of
492 landslide susceptibility from classes 1 to 5 are assigned from low to high according
493 to FS values (e.g. Huang et al., 2007): Class 1 ($FS \leq 1.0$), Class 2 ($1.0 < FS < 1.2$),
494 Class 3 ($1.2 < FS < 1.5$), Class 4 ($1.5 < FS < 2.0$), Class 5 ($FS \geq 2$).

495

496 **4 Conclusions**

497

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

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

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

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

536 **ACKNOWLEDGMENTS**

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

543

544

545

546

547 **Acronyms table**

548

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

549

550

551

552

553

554

555

556

557 **REFERENCES**

558

559 Abera W., A. Antonello, S. Franceschi, G. Formetta, R Rigon , "The uDig Spatial

560 Toolbox for hydro-geomorphic analysis" in GEOMORPHOLOGICAL

- 561 TECHNIQUES, v. 4, n. 1 (2014), p. 1-19. - URL:
562 http://www.geomorphology.org.uk/sites/default/files/geom_tech_chapters/2.4.1_GI
563 SToolbox.pdf
- 564 Beguería, S. (2006). Validation and evaluation of predictive models in hazard
565 assessment and risk management. *Natural Hazards*, 37(3), 315-329.
- 566 Bennett ND, Croke BF, Guariso G, Guillaume JH, Hamilton SH, Jakeman AJ,
567 Marsili-Libelli S, Newham LT, Norton JP, Perrin C, Pierce SA. Characterising
568 performance of environmental models. *Environmental Modelling & Software*. 2013
569 Feb 28;40:1-20.
- 570 Borga, M., Dalla Fontana, G., & Cazorzi, F. (2002). Analysis of topographic and
571 climatic control on rainfall-triggered shallow landsliding using a quasi-dynamic
572 wetness index. *Journal of Hydrology*, 268(1), 56-71.
- 573 Brabb, E.E., (1984). Innovative approaches to landslide hazard and risk mapping,
574 Proceedings of the 4th International Symposium on Landslides, 16–21 September,
575 Toronto, Ontario, Canada (Canadian Geotechnical Society, Toronto, Ontario,
576 Canada), 1:307–324
- 577 Brenning, A. "Spatial prediction models for landslide hazards: review,
578 comparison and evaluation." *Natural Hazards and Earth System Science* 5,
579 no. 6 (2005): 853-862.
- 580 Capparelli, G., & Versace, P. (2011). FLAIIR and SUSHI: two mathematical models
581 for early warning of landslides induced by rainfall. *Landslides*, 8(1), 67-79.
- 582 Capparelli G, Iaquinia P, Iovine GGR, Terranova OG, Versace P. Modelling the
583 rainfall-induced mobilization of a large slope movement in northern Calabria.
584 *Natural Hazards* 2012 ;61:247–256.
- 585 Carrara, A., Merenda, L., 1976. Landslide inventory in Northern Calabria, Southern
586 Italy. *Geological Society of America Bulletin* 87, 1153–1162
- 587 Casadei, M., Dietrich, W. E., & Miller, N. L. (2003). Testing a model for predicting the
588 timing and location of shallow landslide initiation in soil-mantled landscapes. *Earth*
589 *Surface Processes and Landforms*, 28(9), 925-950.
- 590 Cascini, L., Bonnard, C., Corominas, J., Jibson, R., & Montero-Olarte, J. (2005).
591 Landslide hazard and risk zoning for urban planning and development. *Landslide*
592 *Risk Management*. Taylor and Francis, London, 199-235.

- 593 Catani, F., Casagli, N., Ermini, L., Righini, G., & Menduni, G. (2005). Landslide
594 hazard and risk mapping at catchment scale in the Arno River basin. *Landslides*,
595 2(4), 329-342.
- 596 Chung C-JF, Fabbri AG and van Westen CJ (1995) Multivariate regression analysis
597 for landslide hazard zonation. Carrara A and Guzzetti F (Eds.) Geographical
598 Information Systems in assessing natural hazards. Dordrecht, Kluwer Academic
599 Publishers. 5:107-34
- 600 Colella A, De Boer PL, Nio SD. Sedimentology of a marine intermontane Pleistocene
601 Gilbert-type fan-delta complex in the Crati Basin, Calabria, southern Italy.
602 *Sedimentology* 1987;34:721–736.
- 603 Conforti, M., Pascale, S., Robustelli, G., & Sdao, F. (2014). Evaluation of prediction
604 capability of the artificial neural networks for mapping landslide susceptibility in the
605 Turbolo River catchment (northern Calabria, Italy). *Catena*, 113, 236-250.
- 606 Conforti M, Aucelli PPC, Robustelli G, Scarciglia F. Geomorphology and GIS
607 analysis for mapping gully erosion susceptibility in the Turbolo Stream catchment
608 (Northern Calabria, Italy). *Natural Hazards* 2011;56:881–898.
- 609 Corominas J, Van Westen C, Frattini P, Cascini L, Malet JP, Fotopoulou S, Catani F,
610 Van Den Eeckhaut M, Mavrouli O, Agliardi F, Pitilakis K. Recommendations for the
611 quantitative analysis of landslide risk. *Bulletin of engineering geology and the*
612 *environment*. 2014 May 1;73(2):209-63.
- 613 Dietrich, W. E., Bellugi, D. and Real De Asua, R. (2001) Validation of the Shallow
614 Landslide Model, SHALSTAB, for Forest Management, in *Land Use and*
615 *Watersheds: Human Influence on Hydrology and Geomorphology in Urban and*
616 *Forest Areas* (eds M. S. Wigmosta and S. J. Burges), American Geophysical
617 Union, Washington, D. C.. doi: 10.1029/WS002p0195
- 618 David, O., Ascough II, J. C., Lloyd, W., Green, T. R., Rojas, K. W., Leavesley, G. H.,
619 & Ahuja, L. R. (2013). A software engineering perspective on environmental
620 modeling framework design: The Object Modeling System. *Environmental*
621 *Modelling & Software*, 39, 201-213.
- 622 Duan, Q., Sorooshian S., and Gupta V(1992): Effective and efficient global
623 optimization for conceptual rainfall-runoff models. *Water Resources Research* 28.4
624 (1992): 1015-1031.

- 625 Duncan, J. M., and S. G. Wright (2005), Soil Strength and Slope Stability, 297 pp.,
626 New Jersey, John Wiley.
- 627 Fabbricatore D, Robustelli G, Muto F. Facies analysis and depositional architecture
628 of shelf-type deltas in the Crati Basin (Calabrian Arc, south Italy). *Boll. Soc. Geol.*
629 *It.* 2014;133(1):131-148.
- 630 Formetta, G., Mantilla, R., Franceschi, S., Antonello, A., & Rigon, R. (2011). The
631 JGrass-NewAge system for forecasting and managing the hydrological budgets at
632 the basin scale: models of flow generation and propagation/routing. *Geoscientific*
633 *Model Development*, 4(4), 943-955.
- 634 Formetta, G., Antonello, A., Franceschi, S., David, O., & Rigon, R. (2014).
635 Hydrological modelling with components: A GIS-based open-source framework.
636 *Environmental Modelling & Software*, 55, 190-200.
- 637 Formetta, G., Capparelli, G., Rigon, R., and Versace, P.: Physically based landslide
638 susceptibility models with different degree of complexity: calibration and
639 verification. *International Environmental Modelling and Software Society (iEMSS).*
640 *7th Intl. Congress on Env. Modelling and Software*, San Diego, CA, June 15-19,
641 USA, Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.), 2014.
642 http://www.iemss.org/sites/iemss2014/papers/iemss2014_submission_157.pdf
- 643 G. Formetta, G. Capparelli, and P. Versace , Modelling rainfall induced shallow
644 landslides in the Landslide Early Warning Integrated System project Slopes and
645 Geohazards. January 2015, 1747-1752. Available at:
646 <http://www.icevirtuallibrary.com/doi/abs/10.1680/ecsmge.60678.vol4.260>
- 647 Formetta, G., Simoni, S., Godt, J. W., Lu, N., & Rigon, R. (2016). Geomorphological
648 control on variably saturated hillslope hydrology and slope instability. *Water*
649 *Resources Research*.
- 650 Frattini, P., Crosta, G., & Carrara, A. (2010). Techniques for evaluating the
651 performance of landslide susceptibility models. *Engineering geology*, 111(1), 62-
652 72.
- 653 Guzzetti, Fausto, Alberto Carrara, Mauro Cardinali, and Paola Reichenbach.
654 "Landslide hazard evaluation: a review of current techniques and their
655 application in a multi-scale study, Central Italy." *Geomorphology* 31, no. 1
656 (1999): 181-216.

- 657 Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., & Galli, M. (2006).
658 Estimating the quality of landslide susceptibility models. *Geomorphology*, 81(1),
659 166-184.
- 660 Glade, T., & Crozier, M. J. (2005). A review of scale dependency in landslide hazard
661 and risk analysis. *Landslide hazard and risk*, Vol. 3, 75-138.
- 662 Goodenough, D.J., Rossmann, K., Lusted, L.B., 1974. Radiographic applications of
663 receiver operating characteristic (ROC) analysis. *Radiology* 110, 89–95.
- 664 Graham J (1984) Methods of slope stability analysis. In: Brunsten D, Prior DB (eds)
665 Slope instability. Wiley, New York, pp 171–215
- 666 Hay, L.E., G.H. Leavesley, M.P. Clark, S.L. Markstrom, R.J. Viger, and M. Umemoto
667 (2006). Step-Wise, Multiple-Objective Calibration of a Hydrologic Model for a
668 Snowmelt-Dominated Basin. *Journal of the American Water Resources*
669 *Association* 42:877-890, 2006
- 670 Huang, J. C., Kao, S. J., Hsu, M. L., & Liu, Y. A. (2007). Influence of Specific
671 Contributing Area algorithms on slope failure prediction in landslide modeling.
672 *Natural Hazards and Earth System Science*, 7(6), 781-792.
- 673 Lanzafame G, Tortorici L. La tettonica recente del Fiume Crati (Calabria). *Geografia*
674 *Fisica e Dinamica Quaternaria* 1984; 4:11-21.
- 675 Iovine, G., Petrucci, O., Rizzo, V., Tansi, C., 2006. The March 7th 2005 Cavallerizzo
676 (Cerzeto) landslide in Calabria—Southern Italy. *Engineering geology for*
677 *tomorrow's cities—the 10th IAEG congress, Nottingham (UK), The Geological*
678 *Society of London, Paper number 785.*
- 679 Lee, S., Choi, J., & Min, K. (2002). Landslide susceptibility analysis and verification
680 using the Bayesian probability model. *Environmental Geology*, 43(1-2), 120-131.
- 681 Young J, Colella A. Calcareous nannofossils from the Crati Basin. In: Colella A.
682 (ed.), *Fan Deltas-Excursion Guidebook*. Università della Calabria, Cosenza, Italy.
683 79-96; 1988.
- 684 Kennedy, J., and Eberhart R.(1995): Particle swarm optimization. *Neural Networks,*
685 *1995. Proceedings., IEEE International Conference on.* Vol. 4. Perth, WA. IEEE,
686 1995.
- 687 Iovine GGR, Lollino P, Gariano SL, Terranova OG. Coupling limit equilibrium
688 analyses and real-time monitoring to refine a landslide surveillance system in

- 689 Calabria (southern Italy). *Natural Hazards and Earth System Sciences* 2010;
690 10:2341–2354.
- 691 Iverson RM. 2000. Landslide triggering by rain infiltration. *Water Resources*
692 *Research* 36(7): 1897–1910
- 693 Jolliffe, I. T., & Stephenson, D. B. (Eds.). (2012). *Forecast verification: a*
694 *practitioner's guide in atmospheric science*. University of Exeter, UK.
695 John Wiley & Sons.
- 696 Lu, N., and J. Godt (2008), Infinite slope stability under steady unsaturated seepage
697 conditions, *Water Resour. Res.*, 44, W11404, doi:10.1029/2008WR006976.
- 698 Milledge, D. G., Bellugi, D., McKean, J. A., Densmore, A. L., & Dietrich, W. E.
699 (2014). A multidimensional stability model for predicting shallow landslide size and
700 shape across landscapes. *Journal of Geophysical Research: Earth Surface*,
701 119(11), 2481-2504.
- 702 Montgomery, D. R., & Dietrich, W. E. (1994). A physically based model for the
703 topographic control on shallow landsliding. *Water resources research*, 30(4), 1153-
704 1171.
- 705 Murdoch, D. J., & Chow, E. D. (1996). A graphical display of large correlation
706 matrices. *The American Statistician*, 50(2), 178-180.
- 707 Naranjo, J.L., van Westen, C.J. and Soeters, R. (1994) Evaluating the use of training
708 areas in bivariate statistical landslide hazard analysis: a case study in Colombia.
709 *ITC Journal*, 3:292-300.
- 710 Pepe, M.S., 2003. *The Statistical Evaluation of Medical Tests for Classification and*
711 *Prediction*. Oxford University Press, New York.
- 712 Park, N. W. (2011). Application of Dempster-Shafer theory of evidence to GIS-
713 based landslide susceptibility analysis. *Environmental Earth Sciences*,
714 62(2), 367-376.
- 715 Park, H. J., Lee, J. H., & Woo, I. (2013). Assessment of rainfall-induced shallow
716 landslide susceptibility using a GIS-based probabilistic approach. *Engineering*
717 *Geology*, 161, 1-15.
- 718 Pradhan, B. (2011). An assessment of the use of an advanced neural network model
719 with five different training strategies for the preparation of landslide susceptibility
720 maps. *Journal of Data Science*, 9(1), 65-81.

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

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

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

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

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

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

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

740
741
742
743
744
745
746
747

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

750

Name	Definition	Range	Optimal value
------	------------	-------	---------------

Critical success index (CSI)	$CSI = \frac{tp}{tp+fp+fn}$	[0 ,1]	1.0
Equitable success index (ESI)	$ESI = \frac{tp-R}{tp+fp+fn-R}$ $R = \frac{(tp+fn) \cdot (tp+fp)}{tp+fn+fp+tn}$	[-1/3,1]	1.0
Success Index (SI)	$SI = \frac{1}{2} \cdot \left(\frac{tp}{tp+fn} + \frac{tn}{fp+tn} \right)$	[0 ,1]	1.0
Distance to perfect classification (D2PC)	$D2PC = \sqrt{(1-TPR)^2 + FPR^2}$ $TPR = \frac{tp}{tp+fn}$ $FPR = \frac{fp}{fp+tn}$	[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)	$TSS = \frac{(tp \cdot tn) - (fp \cdot fn)}{(tp+fn) \cdot (fp+tn)}$	[-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)	$ACC = \frac{(tp+tn)}{(tp+fn+fp+tn)}$	[0,1]	1.0

751
752
753
754
755
756
757
758
759
760
761
762
763
764

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]

765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790

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).

791

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

792

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

793

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

794

795

796

797

798

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

803

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

804

805

806

807

808

809

810

811

812

813

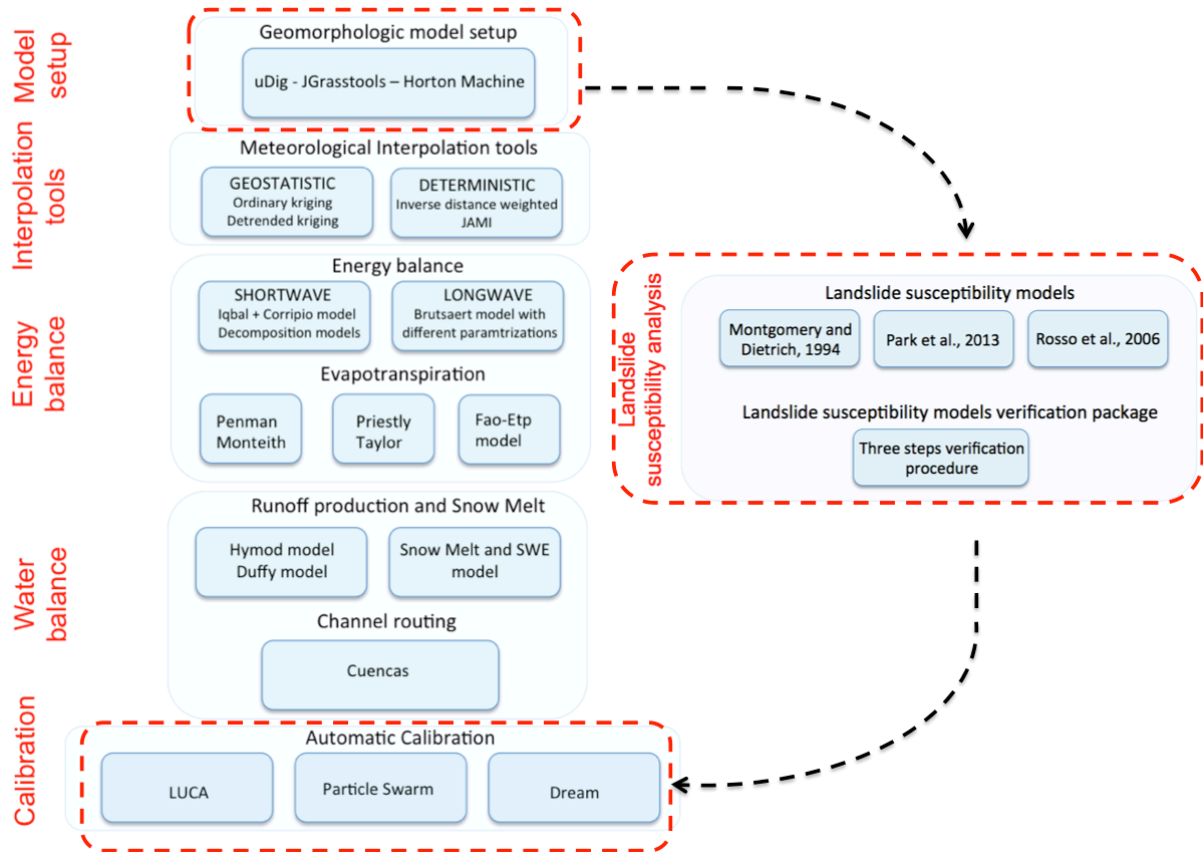
814

815

816

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

817



818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833

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

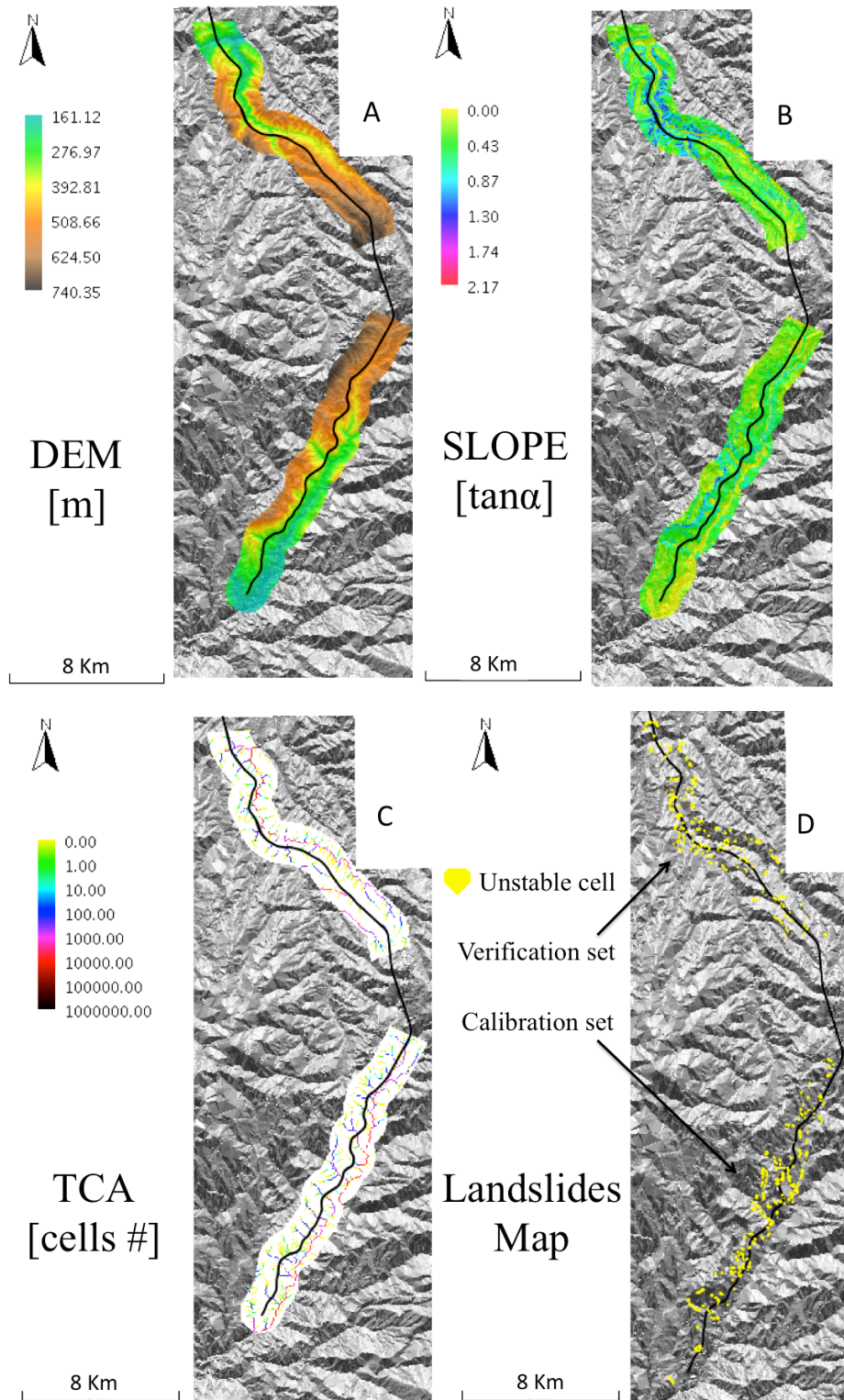


Figure 3: Models' performances results in the ROC plane for M1, M2 and M3. Only GOF indices whose optimization provides $FPR < 0.4$ and $TPR > 0.7$ were reported.

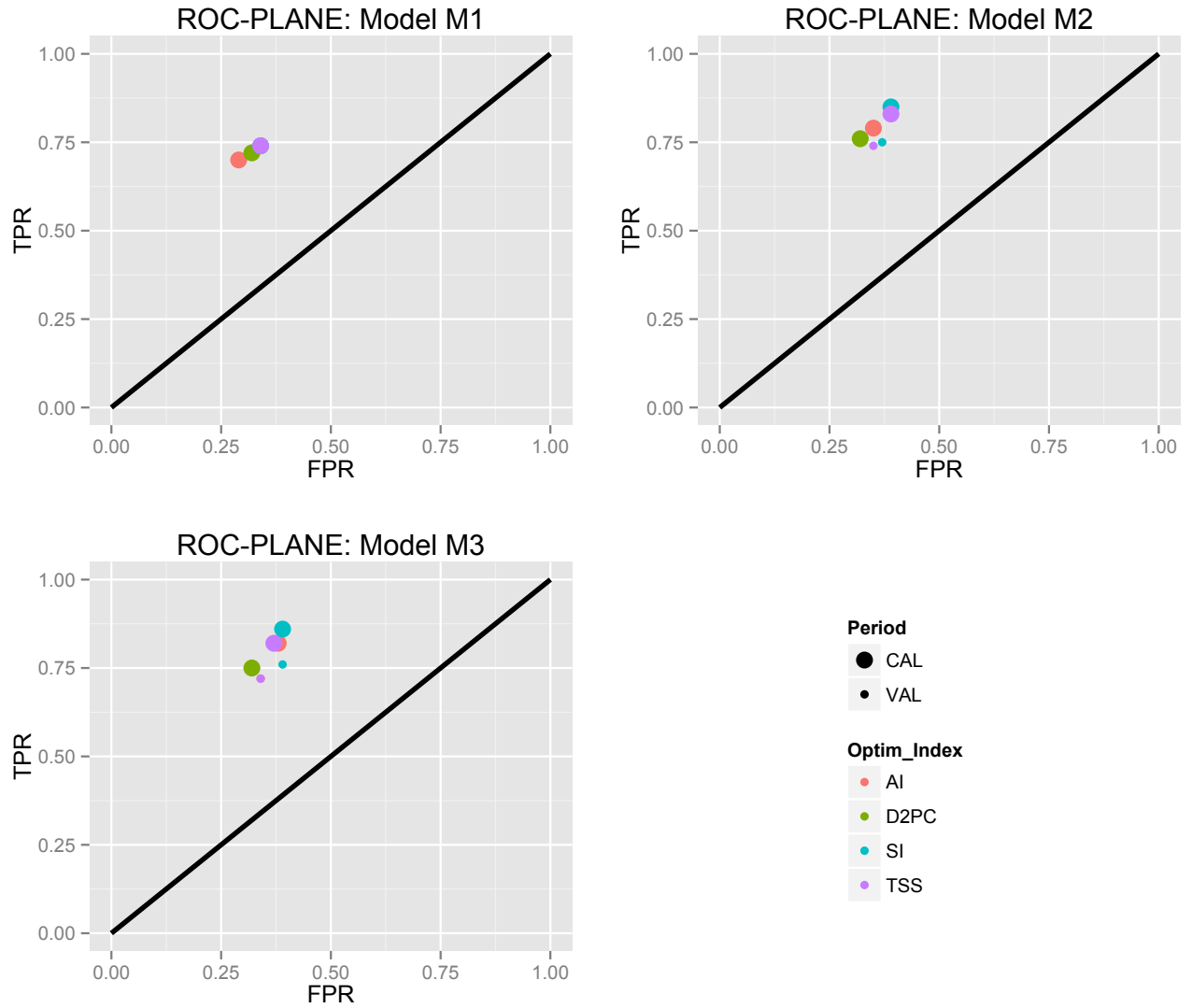


Figure 4: Correlation plot between models' performance (MP) vector computed by optimizing all GOF indices in turn. Results are reported for each model: M1, M2 and M3.

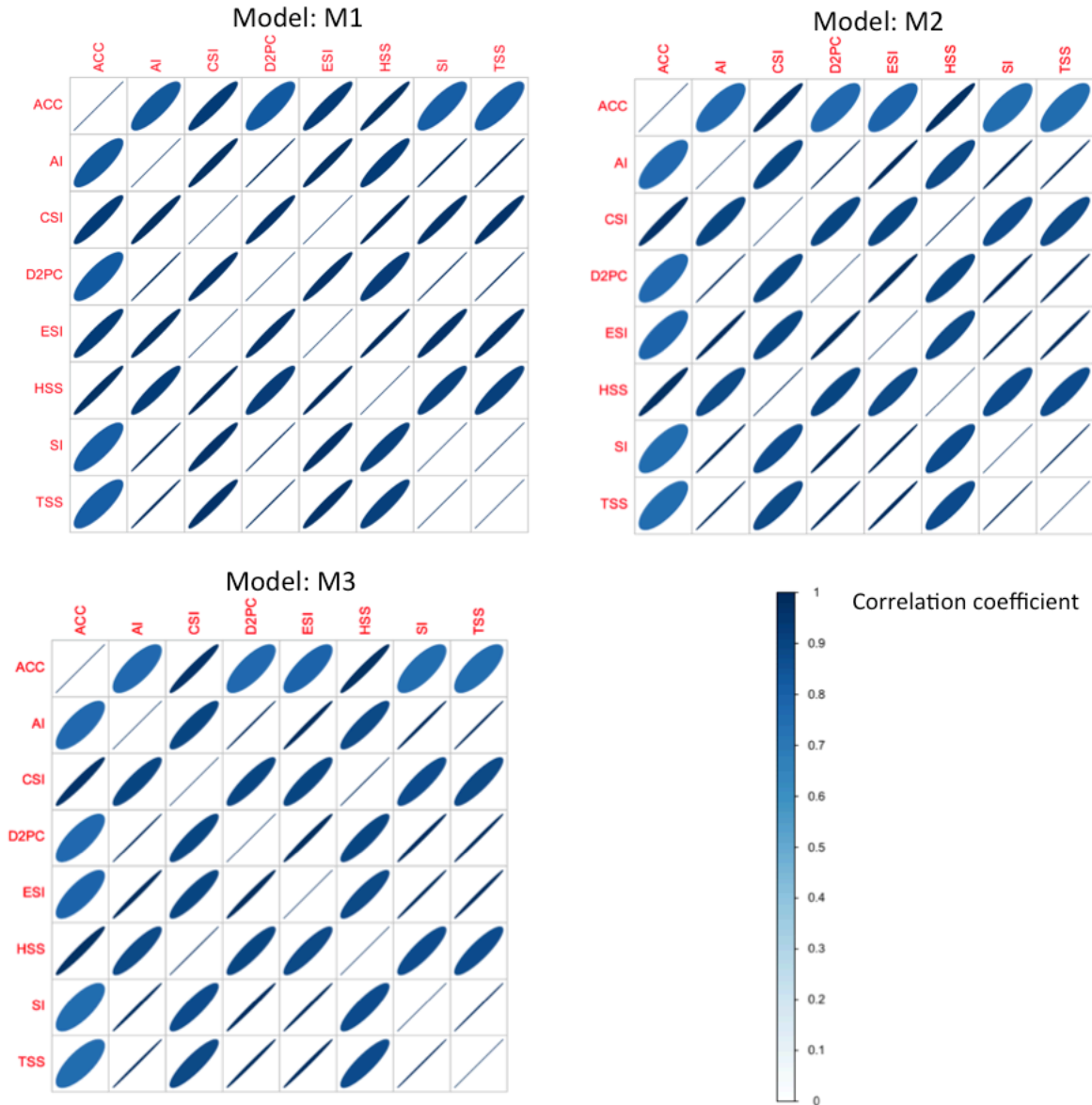


Figure 5: Model M2 parameters sensitivity analysis.



Figure 6: Model M3 parameters sensitivity analysis.

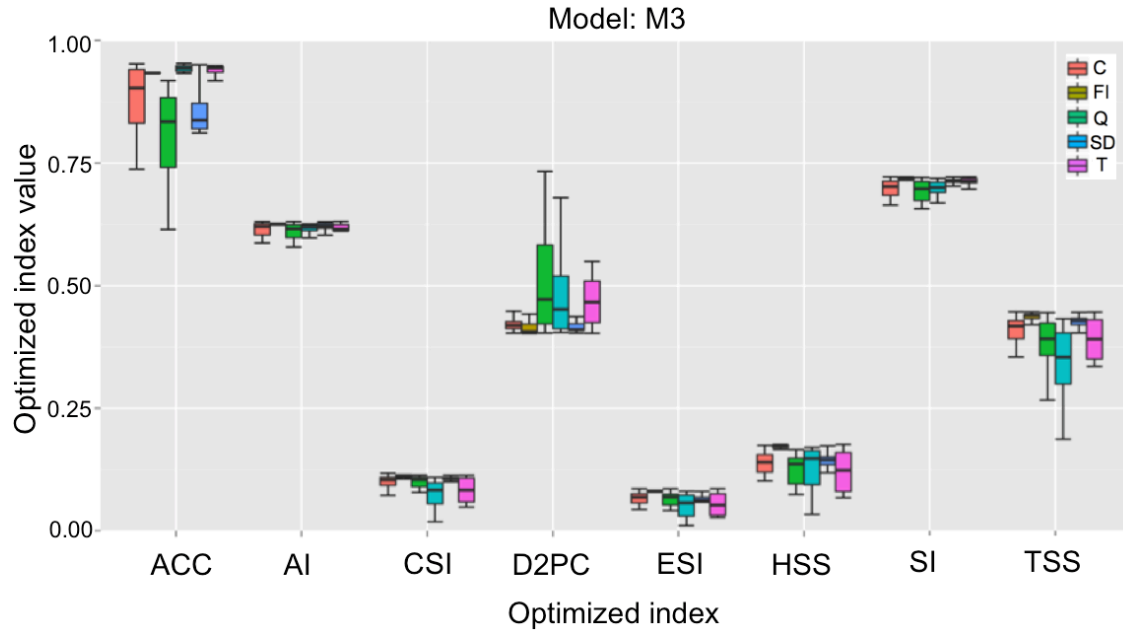
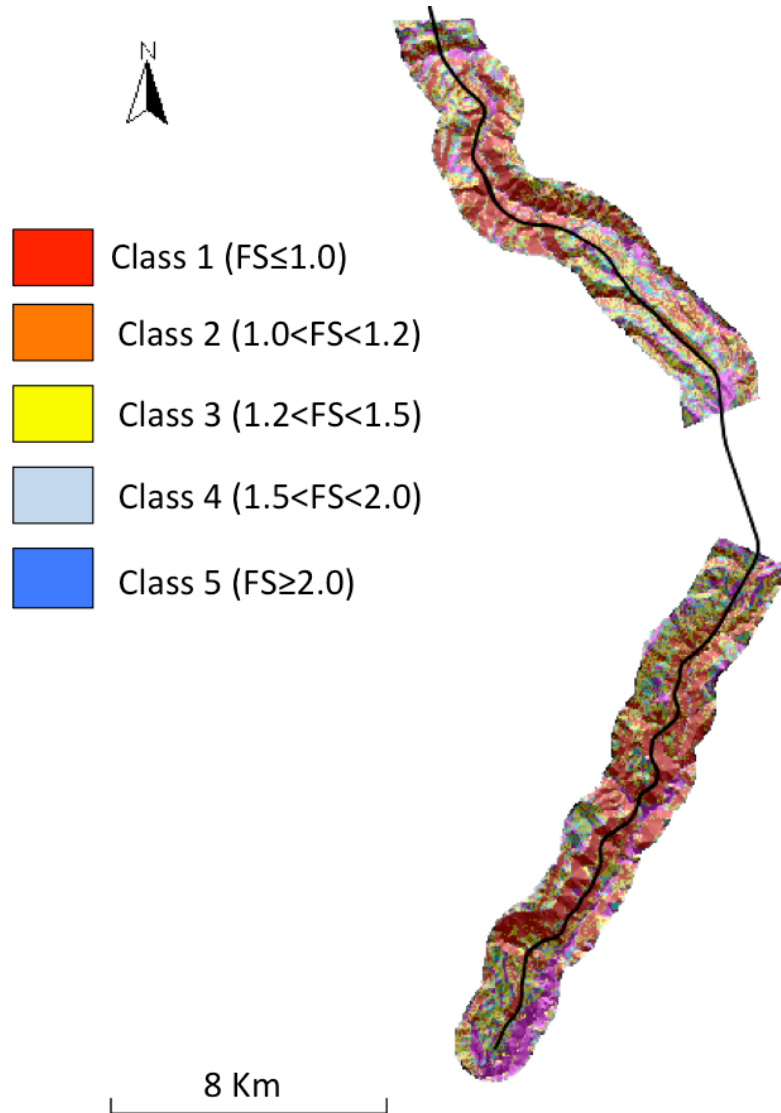


Figure 7: Landslide susceptibility maps using model M3 and parameter set obtained by optimising D2PC.



Appendix 1

1.2 Critical success index (CSI)

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

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

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.

$$ESI = \frac{tp-R}{tp+fp+fn-R} \quad (3)$$

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

1.4 Success index (SI)

SI, eq.(5), equally weights 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.

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

$$TPR = \frac{tp}{tp + fn} \quad (6)$$

$$FPR = \frac{fp}{fp + tn} \quad (7)$$

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.

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

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

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 $M_{opt}=1$.

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

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

The range of the HSS is $-\infty$ 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 t_n is large the false alarm value is relatively overwhelmed. If t_n 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

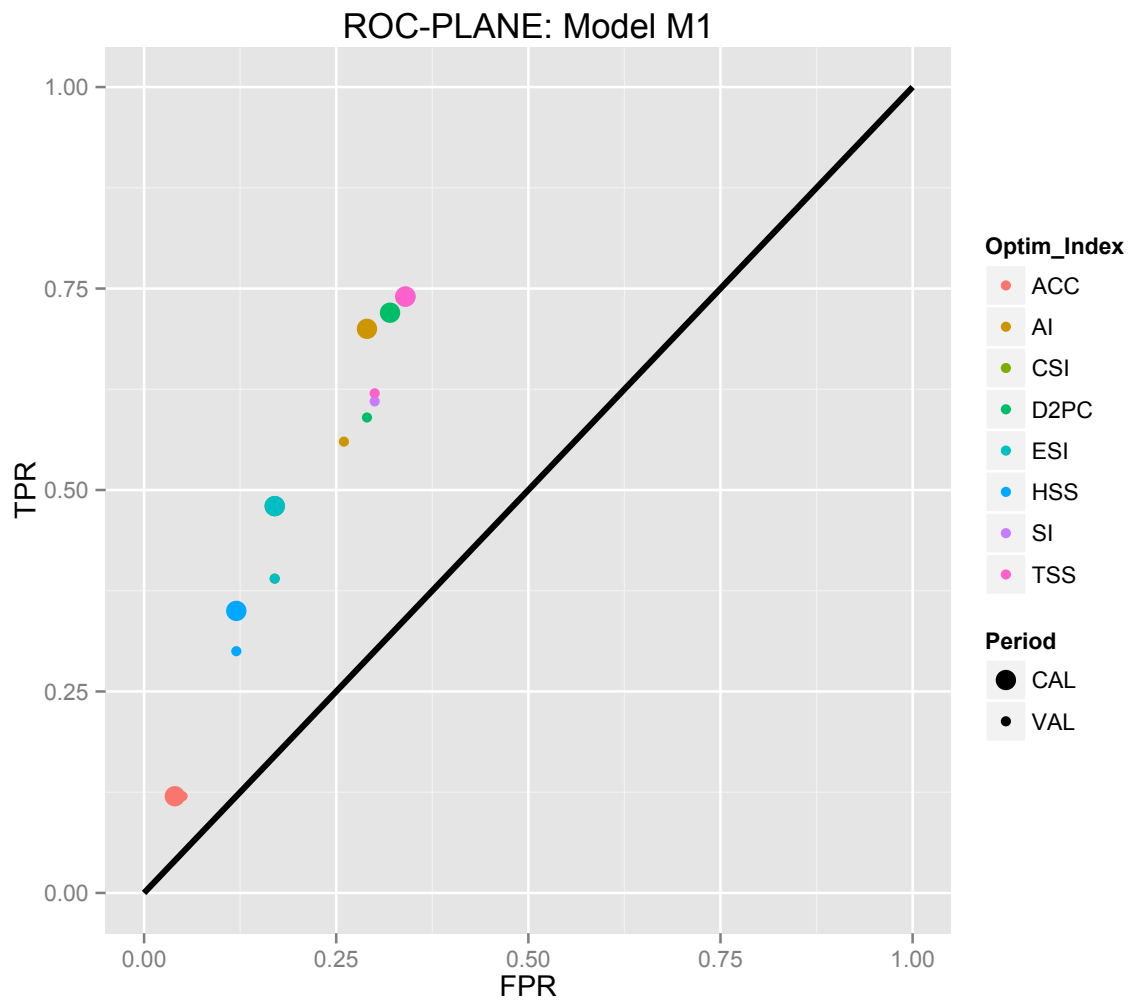


Figure A2-1: Models' performances results in the ROC plane for M1.

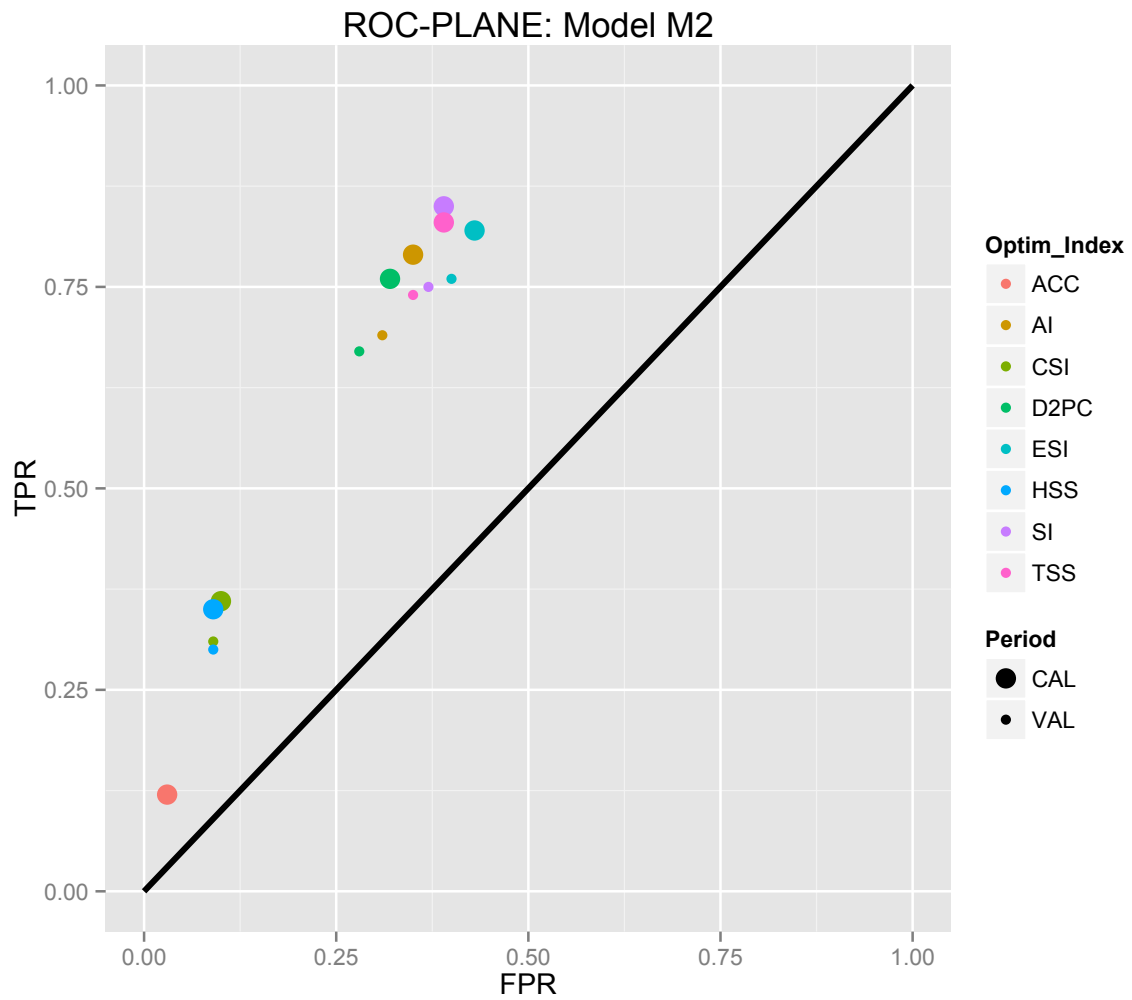


Figure A2-2: Models' performances results in the ROC plane for M2.

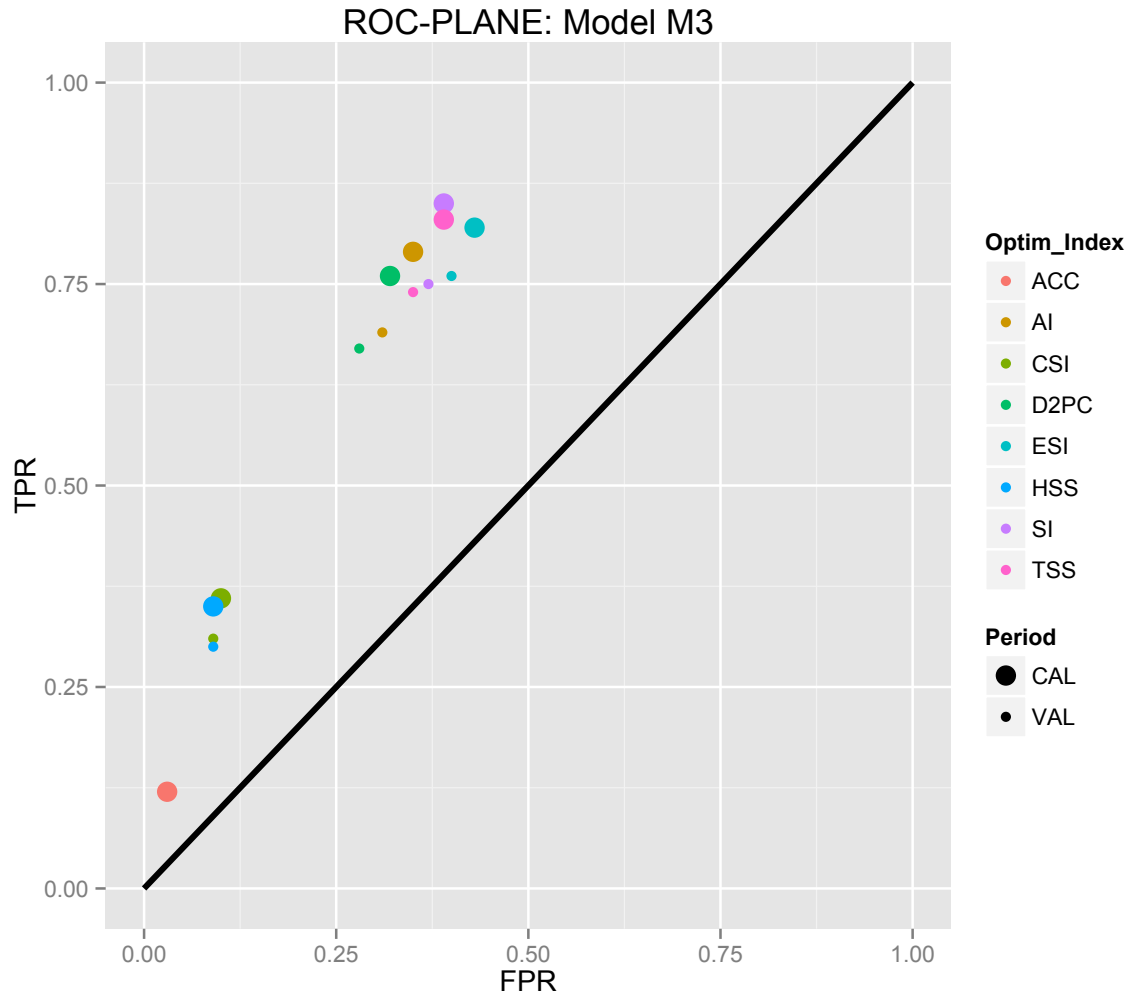


Figure A2-3: Models' performances results in the ROC plane for M3.

