Reply to the Editor comments

“Evaluating performances of simplified physically based models for landslide susceptibility”
G. Formetta, G. Capparelli, P. Versace.

Dear Authors,

please, take into consideration when revising your manuscript all given comments and suggestions by the three reviewers. Especially the suggestions by Reviewer #3 (quoted):

"I would even suggest to rethink the concept and maybe redo the analysis, calibrating only the material parameters. If the data allows, I suggest to use subsets of the landslide inventory which can be assigned to well-defined rainfall events, and to apply the corresponding rainfall intensities and durations to the model." are quite critical.

The Reviewer #1 is more or less easy to incorporate into the revised version of the text. Please, carefully read the text to omit any new misspelled words or typing errors.

After thinking whether to decline the paper or give a free way to proceed with the reviewing process, your answers to the reviewers’ comments show a way out. But, because of some critical comments of the reviewers, the revised version will be sent out for a new round of revision.

Sincerely Yours,

Matjaž Mikoš
Handling Editor
We thank the Editor for his suggestions and comments. We revised the paper according the very useful suggestions of the reviewers and we are happy the reply to reviewers’ comments helped in the revision processes.

After reading the Editor comments, we focused on the question of the reviewer n. 3. We updated the answers to the reviewer n.3 adding new sentences that tried to better take in account of the reviewer’s comment. The file was added in the interactive discussion.

Thanks and best regards

The Authors.
Reply to reviewer n.1: unknown

“Evaluating performances of simplified physically based models for landslide susceptibility”
G. Formetta, G. Capparelli, P. Versace.

We thank the reviewer n. 1 for the revision and the suggestions. We replied in bold below each comment.

Q1) … tool…
A1) We revised the sentence according to the reviewer’s suggestion:
Old sentence: “but also a fundamental tools for the environment”
New sentence: “but also a fundamental tool for the environment”

Q2) Is it 1999 or 2006?
A2) We agree with the reviewer suggestion. The reference Guzzetti et al., 1999 was missing and we added the reference in the revised paper:

Q3) instead “most” use “best”?
A3) We revised the sentence according to the reviewer’s suggestion:
Old sentence: “the choice of the more accurate model”
New sentence: “the choice of the best accurate model”

Q4) reasons
A4) We revised the sentence according to the reviewer’s suggestion:
Old sentence: “For these reason”
New sentence: “For these reasons”
Q5) Brenning is not listed in the References.

A5) We agree with the reviewer suggestion. The reference Brenning, 2005 was missing and we added the reference in the revised paper:


Q6) OMS is a...

A6) We revised the sentence according the reviewer suggestion:

Old sentence: "OMS a Java based modeling framework that promotes"

New sentence: "OMS is a Java based modeling framework that promotes"

Q7) Worku is missing in the References

A7) We agree with the review comment. We had a cited Worku in a wrong way, the correct work is Abera et al 2015 and Abera is currently in the references.

Q8) Rosso et al., 2006

A8) We agree with the review suggestion and we revised twice accordingly:

Old sentence: "Rosso et al 2008"

New sentence: "Rosso et al 2006"

Q9) .. slope gradient ...

A9) We agree with the review suggestion and we revised accordingly:

Old sentence: "slope gradient"

New sentence: "slope gradient, "

Q10) .. slope gradient ...

A10) We agree with the review suggestion and we revised accordingly:

Old sentence: "angle"

New sentence: "angle, "

Q11) .. slope gradient ...

A11) We agree with the review suggestion and we revised accordingly:

Old sentence: "soil"
New sentence: “...soil,”

Q12) Add Worku et al., 2014 to reference list.

A12) We solved the problem of the reference Abera et al 2016 as specified in answer A7.

Q13) Results are presented...

A13) We agree with the reviewer’s suggestion and we revised the sentence:

Old sentence: Results were presented in Table

New sentence: Results are presented in Table

Q14) Provide not provides

A14) We agree with the reviewer’s suggestion and we revised the sentence:

Old sentence: For the model M2 and M3 it is clear that ACC, HSS, and CSI provides the less performing models results

New sentence: For the model M2 and M3 it is clear that ACC, HSS, and CSI provide the less performing models results

Q15) ...are similar to each other...

A15) We agree with the reviewer’s suggestion and we revised the sentence:

Old sentence: ...are similar to each others...

New sentence: ...are similar to each other...

Q16) ...the third step shows

A16) We agree with the reviewer’s suggestion and we revised the sentence:

Old sentence: ... the third step show

New sentence: ... the third step shows

Q17) ... fact accommodate

A17) We agree with the reviewer’s suggestion and we revised the sentence:

Old sentence: A more sensitive couple model-optimal parameter set will in fact accommodates

New sentence: A more sensitive couple model-optimal parameter set will in fact
accommodate

Q18) … according to FS
A18) We agree with the reviewer’s suggestion and we revised the sentence:
Old sentence: are assigned from low to high according FS
New sentence: are assigned from low to high according to FS

Q19) … this allows the...
A19) We agree with the reviewer’s suggestion and we revised the sentence:
Old sentence: and this allow the user to
New sentence: and this allows the user to

Q20) … this allows the...
A20) We agree with the reviewer’s suggestion and we revised the sentence:
Old sentence: is the number of correct detected landslide pixels
New sentence: is the number of correct detected landslide pixels

Q21) … measures...
A21) We agree with the reviewer’s suggestion and we revised the sentence:
Old sentence: It measure the distance
New sentence: It measures the distance

Q22) performance with respect
A22) We agree with the reviewer’s suggestion and we revised the sentence:
Old sentence: to quantify the model performance respect to set of control or reference model
New sentence: to quantify the model performance with respect to set of control or reference model

Q23) delete "that indicates"
A23) We agree with the reviewer’s suggestion and we revised the sentence:
Old sentence: Negative values indicate that indicates that the mod
Q24) treats

A24) We agree with the reviewer’s suggestion and we revised the sentence:

Old sentence: A problem of TSS is that it threatens the hit rate
New sentence: A problem of TSS is that it treats the hit rate

Q25) This reference is not mentioned in the text.

A25) We removed the reference:


Q26) This reference is not mentioned in the text.

A26) We removed the reference:


Q27) This reference is not mentioned in the text.

A27) We did not remove the reference Fabbricatore et al., 2014 because is in the sentence:

“The Crati Basin is a Pleistocene-Holocene extensional basin filled by clastic marine and fluvial deposits (Vezzani, 1968, Colella et al., 1987, Fabbricatore et al., 2014).”

Q28) This reference is not mentioned in the text.

A28) We do not deleted the reference Formetta et al., 2015 because is in the text but was indicated as Formetta et al. 2014. So we fixed the error:

Old sentence: The landslide susceptibility models implemented in NewAge-JGrass and presented in a preliminary application in Formetta et al., 2014
New sentence: The landslide susceptibility models implemented in NewAge-JGrass and presented in a preliminary application in Formetta et al., 2015

Q29) This reference is not mentioned in the text.

A29) We removed the reference:

Q30) This reference is not mentioned in the text.

A30) We did not remove the reference Jolliffe and Stephenson, (2012) because is in the sentence:

"Accurate discussions about the most common quantitative measures of goodness of fit (GOF) between measured and modeled data are available in Bennet et al., (2013), Jolliffe and Stephenson, (2012), Beguería (2006), Brenning (2005) and references therein”

Q31) This reference is not mentioned in the text.

A31) We removed the reference:


Q32) This reference is not mentioned in the text.

A32) We removed the reference:


Q33) This reference is not mentioned in the text.

A33) We removed the reference:


Q34) This reference is not mentioned in the text.

A34) We removed the reference:

Q35) Results are presented...
A35) We agree with the reviewer’s suggestion and we revised the sentence:
Old sentence: Results were presented for each model
New sentence: Results are presented for each model

Q36) calibration (CAL) and verification (VAL).
A36) We agree with the reviewer’s suggestion and we revised the sentence:
Old sentence: calibration and verification.
New sentence: calibration (CAL) and verification (VAL).

Q37) are shown
A37) We agree with the reviewer’s suggestion and we revised the sentence:
Old sentence: In bold the rows for which
New sentence: In bold are shown the rows for which

Q38) GIS is written twice and Geographic is missing a letter “a”.
A38) We removed one of the GIS and we fixed the typo:
Old sentence: Geogrphic informatic system
New sentence: Geographic informatic system

Q39) The text is small and consequentially hard to read.
A39) We revised the font of the figure according the reviewer’s suggestion
Old version:
New version:
Q40) Could you scale up the section where the scores are shown to emphasise the differences?

A40) We thank the author for the suggestion but we prefer to maintain the complete dimension of the ROC space, this will help the reader to easily understand the differences between the three models. Moreover a full representation of all the models is reported in appendix.

Q41) The text is small and consequentially hard to read.

A41) We revised the font of the figure according the reviewer’s suggestion

Old version:

New version:
Q42) The text is small and consequently hard to read.

A42) We revised the font of the figure according to the reviewer’s suggestion

Old version:

New version:
Q43) What is the meaning of classes 1-5? I suggest you put the values of FS with the class tags (Class 1 (FS< 1.0), Class 2 (1.0 <FS< 1.2), Class 3 (1.2 <FS< 1.5), Class 4 (1.5 <FS< 2.0), Class 5 (FS> 2))

A43) We agree with the reviewer’s suggestion and we modified the figure accordingly:

Old version:
New version:
Formetta et al. / Evaluating performances of simplified physically based landslide susceptibility models
Reply to reviewer n.2: unknown

“Evaluating performances of simplified physically based models for landslide susceptibility”
G. Formetta, G. Capparelli, P. Versace.

We thank the reviewer n. 2 for the revision and the suggestions. We replied in bold below each comment.

GENERAL COMMENTS

This manuscript (MS) presents an interesting and important topic on GIS-based landslides susceptibility mapping. However, the MS has some flaws that need to be taken care of.

Q1) Geology, hydrogeology and land cover are important factors in landslide susceptibility study. As mention in the Abstract of this MS, the authors only mentioned “hydrology, geotechnical science, geomorphology, and statistics.”

A1) We agree with the reviewer’s comment and we revised the sentence in the abstract adding geology and hydrogeology as important factors in landslide susceptibility analysis:

Old sentence: “Prediction of shallow landslides susceptible locations is a complex task that involves many disciplines: hydrology, geotechnical science, geomorphology, and statistics”.

New sentence: “Prediction of shallow landslides susceptible locations is a complex task that involves many disciplines: hydrology, geotechnical science, geology, hydrogeology, geomorphology, and statistics”.

Moreover in the introduction we took into account of the importance of geology on landslide susceptibility. Specifically in the sentence: “Geo-environmental factors such as geology, land-use, vegetation, climate, increasing population may increase the landslides occurrence (Sidle and Ochiai 2006).”
Q2) The MS has never mentioned the types of landslide (or failure mechanisms), e.g. translational or rotational landslide that they were modeling. It is important to identify the landslide type first and then select the proper physical model.

A2) We agree with the reviewer’s suggestion and we added the following sentence to specify for what kind of failure mechanism the models are more suitable. Moreover the new sentence answer also to the Q3 reviewer comment where is asked to define what a shallow landslide is:

New sentence: “Those models are suitable for shallow translational landslides controlled by groundwater flow convergence. Shallow landslides usually have a very low ratio between the maximum depth (D) and the length (L) of scar (D/L<0.1, Casadei et al., 2003); involve small volume of the colluvial soil mantle and present a generally translational failure mechanism (Milledge et al., 2014)”

Q3) The MS keeps referring to “shallow landslide”. What is the definition of “shallow landslides”? What is the failure mechanism of a “shallow landslide”?

A3) We hope that in the answer A2 we have meet this reviewer request.

Q4) There are so many grammar errors and typos, which distract me from reading the MS. I list examples of these errors and typos under “Suggested Edits”. I don’t think I found all of them. I strongly suggest that the authors should have someone editing their writing carefully in order to make this MS publishable.

A4) We revised all the grammar error suggested by the reviewer 2. Moreover, we revised again the language and the typos in the paper taking into account the typos that also the reviewer 1 pointed out.

SPECIFIC COMMENTS

Here is a list of additional items need to be addressed:

Q5) As stated in the MS

“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.”
However, degree of soil saturation could either be a stabilizing or destabilizing factor depends on the geomorphology, e.g. slope angle. 2

A5) We agree with the reviewer’s suggestion. In the sentence we wanted to point out two features of the model M2: 1) the fact that consider the effect of the degree of soil saturation and void ratio above the groundwater table and ii) the fact that consider the stabilizing effect of the soil cohesion. We revised the sentence according to the reviewer’s suggestion:

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

Q6) Equation (3) – the meanings of symbols need to be explained.

A6) We partially agree with the reviewer’s comment. There were only two symbols in eq. 3 that were not explained: degree of saturation and void ratio. We hope that the sentence that we added in A5, were we specify the symbols $S_r$ and $e$, has met the reviewer suggestion.

Q7) Appendix A and Table are redundant

A7) We thank the reviewer for the comment but we believe that table are useful to quantify the model performances that sometimes are not easily distinguish in the plot and the appendix A is useful to show the behavior of all the optimized indices in the roc p

SUGGESTED EDITS

Q8) Line 8
a fundamental tools

A8) We revised the sentence according the reviewer’s suggestion:

New Sentence: “but also a fundamental tool for the environment preservation and a responsible urban planning”

Q9) Line 10
During the last decades
Or
During the last few decades

A9) We revised the sentence according to the reviewer’s suggestion:
New sentence: “During the last few decades many methods for landslide susceptibility mapping”

Q10) Lines 18-19
to link instability factors (such as geology, soils, slope, curvature, and aspect) and past and present landslides.

A10) We revised the sentence according to the reviewer’s suggestion:
New sentence: “Use different approaches such as multivariate analysis, discriminant analysis, random forest to link instability factors (such as geology, soils, slope, curvature, and aspect) with the past and present landslides.”

Q11) Lines 24-25
The soil-stability component simulates the safety factor of the slope safety factor (FS) defined as ratio of stabilizing to destabilizing forces.

A11) We revised the sentence according to the reviewer’s suggestion:
New sentence: “The soil-stability component simulates the slope safety factor (FS) defined as ratio of stabilizing to destabilizing forces.”

Q12) Line 5
For these reasons,

A12) We revised the sentence according to the reviewer’s suggestion:
New sentence: “For these reasons,”
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.

We thank the reviewer for the suggestion we modified the sentence using an “and” between open-source and GIS based because they both are adjectives of hydrological model. The new sentence is:

New Sentence: "The procedure is implemented in the open source and 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 the model developers with many facilitates such as: multithreading, implicit parallelism, models interconnection, GIS based system.

We revised the sentence according the reviewer’s suggestion:

New sentence: OMS is a Java based modeling framework that promotes the idea of programming by components and provides the model developers with many facilitates such as: multithreading, implicit parallelism, models interconnection, and GIS based system.
Comparing the results obtained for different models and for different GOF metrics, the user can select the most performing combination for the own case study. 

Or

Comparing the results obtained for different models and for different GOF metrics, the user can select the most performing combination for his or her own case study.

Q16) Lines 19-21

Thus different LSA configurations can be realized depending on: the landslide susceptibility model, the calibration algorithm, and the GOFs selected by the user. 

Q17) Lines 24-26

the Montgomery and Dietrich (1994) model (M1), the Park et al. (2013) model (M3) and the Rosso et al. (2008) model (M3).

the Montgomery and Dietrich (1994) model (M1), the Park et al. (2013) model (M2) and the Rosso et al. (2006) model (M3).

Q18) Line 5

a [-] is the slope gradient
a \( \alpha [-] \) is the slope angle

**A18** We revised the sentence according the reviewer’s suggestion:

New sentence: “\( \alpha [-] \) is the slope angle”

**Q19** Lines 12-13

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.

In order to assess the models’ performance we developed a model that computes
the most used indices for assessing the quality of a landslide susceptibility map.

**A19** We revised the sentence according the reviewer’s suggestion:

New sentence: In order to assess the models’ performance we developed a model that computes
the most used indices for assessing the quality of a landslide susceptibility map.

**Q20** Lines 16-17

This is possible because each model is an OMS component and can be linked to the
 calibration algorithms as it is, without rewriting or modifying their code.

This is possible because each model is an OMS component and can be linked to the
calibration algorithms as it is, without rewriting or modifying its code.

**A20** We revised the sentence according the reviewer’s suggestion:

New sentence: “This is possible because each model is an OMS component and can
be linked to the calibration algorithms as it is, without rewriting or modifying its code”.

**Q21** Lines 7-8

Secondly, we verified if each OF metric has own information content or if it provides
information analogous to other metrics (and unessential).

Secondly, we verified if each OF metric has its own information content or if it
provides information analogous to other metrics (and unessential).

**A21** We revised the sentence according the reviewer’s suggestion:
New sentence: "Secondly, we verified if each OF metric has its own information content or if it provides information analogous to other metrics (and unessential)."

Q22) Lines 1-2
Slope gradients, computed from 10m resolution digital elevation model, range from 0 to 55°, while its average is about 26°.

Slope, computed from 10m resolution digital elevation model, ranges from 0° to 55°, with its average is about 26°.

A22) We revised the sentence according the reviewer’s suggestion:
New sentence: “Slope, computed from 10 meters resolution digital elevation model, range from 0° to 55°, while its average is about 26°.”

Q23) Lines 7-9
The first unit is a Lower Pliocene succession of conglomerates and sandstones passing upward into silty clays (Lanzafame and Tortorici, 1986) second unit.

The first unit is a Lower Pliocene succession of conglomerates and sandstones passing upward into the silty clays (Lanzafame and Tortorici, 1986) second unit.

A23) We revised the sentence according the reviewer’s suggestion:
New sentence: “The first unit is a Lower Pliocene succession of conglomerates and sandstones passing upward into silty clays (Lanzafame and Tortorici, 1986) second unit.”

Q24) Lines 11-12
as also suggested by data provided by Young and Colella, 1988.

as also suggested by data provided by Young and Colella (1988).

A24: We revised the sentence according the reviewer’s suggestion:
New sentence: “as also suggested by data provided by Young and Colella (1988)”

Q25) Lines 15-16
All the data were digitized and stored in GIS database (Conforti et al., 2014) and the results was the map of occurred landslide presented in Fig. 2d.

All the data were digitized and stored in a GIS database (Conforti et al., 2014) and the result was the map of occurred landslide presented in Fig. 2d.

A25) We revised the sentence according the reviewer’s suggestion:
All the data were digitized and stored in a GIS database (Conforti et al., 2014) and the result was the map of occurred landslide presented in Fig. 2d.

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.

For the models M2 and M3 it is clear that ACC, HSS, and CSI provide the less performing models results.
Results presented in Fig. 3 and Table 4 show that:

A29) We revised the sentence according the reviewer’s suggestion:

New sentence: “Results presented in Figure 3 and Table 4 show that:”

Q30) Line 26

for each model M1, M2 and M3.

A30) We revised the sentence according the reviewer’s suggestion:

New sentence: “for each model M1, M2 or M3.”

Q31) Lines 1-2

The more is prominent as the less the vector are correlated;

A31) We revised the sentence according the reviewer’s suggestion:

New sentence: “The more prominent the less the vectors are correlated; ”

Q32) Lines 6-7

This confirms that an optimization of AI, D2PC, SI and TSS provide quite similar model performances,

A32) We revised the sentence according the reviewer’s suggestion:

New sentence: “This confirms that an optimization of AI, D2PC, SI and TSS provides quite similar model performances”

Q33) Line 12

In this step we focused the attention on the models M2 and M3

A33) We revised the sentence according the reviewer’s suggestion:
New sentence: “In this step we focused on the models M2 and M3”

Results where presented in Figs. 5 and 6 for model M2 and M3 respectively.

Results were presented in Figs. 5 and 6 for models M2 and M3 respectively.

A34) We revised the sentence according the reviewer’s suggestion:

New sentence:” Results were presented in Figures 5 and 6 for models 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).

A35) We revised the sentence according the reviewer’s suggestion:

New sentence: “Each column of the figures represents one optimized index and has a number of boxplots 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 to a certain model parameters change.

A36) We revised the sentence according the reviewer’s suggestion:

New sentence: “Each boxplot represents the range of variation of the optimized index due to 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.

The narrower the boxplot for a given optimized index the less sensitive is the model to that parameter.

A37) We revised the sentence according the reviewer’s suggestion:
New sentence: “The narrower the boxplot for a given optimized index the less sensitive is the model to that parameter”
The selection of the more appropriate model for computing landslide susceptibility maps is based on what we learn from the previous steps.

For this reason we used the combination the model M3 with parameters obtained.
"Evaluating performances of simplified physically based models for landslide susceptibility"

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

We thank the reviewer prof. Martin Mergili for the revision and the suggestions. We replied in bold below each comment.

Q1) The paper is interesting and worth publishing in principle. I broadly agree with the comments of Reviewers #1 and #2 but have some additional comments the authors should consider before the manuscript is published.

From a purely technical point of view, the authors present – as far as I can see it – a clear and clean way of parameter calibration/optimization for slope stability modelling.

However, I have some major concerns with regard to the scientific meaningfulness of the approach: while it may be useful to calibrate the material parameters I am not sure how much sense it makes to calibrate such a large number of variables, including the intensity and duration of rainfall. The fact that even the magnitude of the triggering event has to be calibrated means in my opinion that the physically-based model by itself may completely fail to reproduce the processes under investigation, but the input may be tuned in a way that the results somehow fit to the observations. Consequently, the model would have no capability to be applied for making predictions e.g., for a potential future rainfall event of a defined magnitude in the study area. For just mapping the general landslide susceptibility, a comparatively simple and easily reproducible statistical approach would do the work. Consequently, I suggest to at least define more clearly in the introductory chapter what are the specific aims of your study and what you finally intend with this very comprehensive calibration.

Further, this issue has to be addressed appropriately in the discussion.

A1) We thank the reviewer for the comment and we partially agree with it. As concern the approach of model input data calibration (in particular the rainfall) it was used in other studies (e.g. Deb and El-Kadi (2009), Bischetti and
Chiaradia (2010), Huang and Kao (2006)) where the ratio rainfall over soil transmissivity \((R/T)\) was considered uncertain.

As concern the predictive capability of the models we used to test our methodology we fully agree with the reviewer: being the models based on steady state hypothesis they cannot be used for early warning systems or making landslide prediction. We agree with the reviewer we have to specified it better in the text and. We revised the sentence in the introduction section to better clarify that the objective of the paper is not to predict landslide but to test a general methodology for evaluating in a quantitative manner the ability of distributed environmental models in modeling and simulating observed phenomena:

**Old sentence:** “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.”

**New sentence:** “In this work we propose an objective methodology for environmental models analysis that allows to select the most performing model based on a quantitative comparison and assessment of models prediction skills. In this paper the methodology is applied for assessing the performances of simplified landslide susceptibility models. Moreover, being the methodology model independent, it can be used for assessing the ability of any type of environmental model to simulate natural phenomena.”

Q2) Strictly speaking, a landslide inventory should only be used for the evaluation of a coupled hydraulic-slope stability model if it relates to the same triggering event as applied in the modelling (see also comment above!). In general, more information on the landslide inventory should be provided: does it cover only the initiation areas of the landslides, or also the runout zones (in the latter case, it should not be used for evaluating a slope stability model).

A2) We agree with the reviewer comment. We specified in a new sentence in the “Site description” section the fact that the landslide inventory covers only
the initiation area of the landslide and that the used models do not landslide propagation after the triggering:

New sentence: “The landslide inventory map refers only to the initiation area of the landslides. This allows a fair comparison with the landslide models that provide only the triggering point and not include a runout model for landslides propagation.”

In summary, I have the feeling that the authors have done a really fine work in implementing and explaining the computational aspect of their calibration and evaluation procedure. In contrast, they still have to reflect the scientific meaningfulness of the case study employed. At least some aspects should be explained and justified in a clearer way. I would even suggest to rethink the concept and maybe re-do the analysis, calibrating only the material parameters. If the data allows, I suggest to use subsets of the landslide inventory which can be assigned to well-defined rainfall events, and to apply the corresponding rainfall intensities and durations to the model.

A3) We thank the reviewer for the suggestions and we agree in part with it. On one side, we hope that in the answer A1 we were able to better clarify the issue of the calibration of the rainfall input data. It was also performed in other studies and it could be considered meaningful. On the other side we agree with the suggestion of the reviewer and in the conclusion section of the paper we clarify better the aim of the paper (to present and implementing an objective procedure for calibration and evaluation of environmental models). We hope that in the answer 1 we have better clarified that the evaluation of early warning system was not an objective of the paper:

Old sentence: “The paper presents a procedure for landslides susceptibility models evaluation and selection”

New sentence: “The paper presents a procedure quantitatively calibrate, evaluate, and compare the performances of environmental models. The procedure was applied for the analysis of three landslides susceptibility models.”
The authors should feel free to contact me at martin.mergili@univie.ac.at in case they disagree with my comments or if they would like to discuss the one or the other issue.

With best regards, Martin Mergili

References


Reply to reviewer n.3: M. Mergili

“Evaluating performances of simplified physically based models for landslide susceptibility”
G. Formetta, G. Capparelli, P. Versace.

We focused more on one of the question raised by the reviewer n.3 and we added two more sentences in the text regarding the available data in the study area, and the calibration of the steady-state rainfall. The question of the reviewer was:

“In summary, I have the feeling that the authors have done a really fine work in implementing and explaining the computational aspect of their calibration and evaluation procedure. In contrast, they still have to reflect the scientific meaningfulness of the case study employed. At least some aspects should be explained and justified in a clearer way. I would even suggest to rethink the concept and maybe re-do the analysis, calibrating only the material parameters. If the data allows, I suggest to use subsets of the landslide inventory which can be assigned to well-defined rainfall events, and to apply the corresponding rainfall intensities and durations to the model.”

The two new sentences added in conclusion of the revised paper are:

“In the application we presented the effective precipitation was calibrated because we were performing a landslide susceptibility analysis and it was useful for demonstrating the method. However, we are aware that for operational landslide early warning systems the rainfall constitutes a fundamental input of the predictive process”.

“Moreover, the analysis would profit from measured rainfall data that triggered the occurred landslides, but that such data are not available at the moment for the study area".
Evaluating Performances of Simplified Physically Based Models for Landslide Susceptibility.

Giuseppe Formetta, Giovanna Capparelli and Pasquale Versace

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

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

Abstract: Rainfall induced shallow landslides 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, geology, hydrogeology, 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 Altìlia 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.

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

### 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 few 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) with the past and present landslides. Deterministic models (e.g. Montgomery and Dietrich, 1994, Lu and Godt, 2008, Borga et al., 2002, Simoni et al., 2008, Capparelli and Versace, 2011, Lu and Godt, 2013) synthetize 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 slope safety factor (FS) defined as ratio of stabilizing to destabilizing forces.

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 best accurate model, and the selection of the most performing method for parameter estimation are still opened. For these reasons, a procedure that allows 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) and 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 Bennet et al., (2013), Jolliffe and Stephenson, (2012), Beguería (2006), Brenning (2005) and references therein. We summarized them in Appendix 1. 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 environmental models analysis that allows to select the most performing model based on a quantitative comparison and assessment of models prediction skills. In this paper the methodology is applied for assessing the performances of simplified landslide susceptibility models. Moreover, being the methodology model independent, it can be used for assessing the ability of any type of environmental model to simulate natural phenomena. The procedure is implemented in the open source and 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 the model developers with many facilities such as: multithreading, implicit parallelism, models interconnection, and 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 uDig-Spatial Toolbox (Abera et al., (2014), https://code.google.com/p/jgrasstools/wiki/JGrassTools4udig) are used as visualization and input/output data management system.

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

For a generic landslide susceptibility component it is possible to estimate the model parameters that optimize a given GOF metric. To perform this step the user can choose between a set of GOF indices and a set of automatic calibration algorithms.

Comparing the results obtained for different models and for different GOF metrics the user can select the most performing combination for his or her own case study.

The methodology, accurately presented in section 2, was setup considering three different landslide susceptibility models, eight GOF metrics, and one automatic calibration algorithm. The flexibility of the system allows to add more models, GOF metrics, and to use different calibration algorithms. Thus different LSA configurations can be realized depending on: the landslide susceptibility model, the calibration algorithm, and the GOFs selected by the user.

Lastly, section 3 presents a case study of landslide susceptibility mapping along the A3 Salerno-Reggio Calabria highway in Calabria, that illustrates the capability of the system.

2 MODELING FRAMEWORK
The landslide susceptibility analysis (LSA) is implemented in the context of NewAge-JGrass (Formetta et al., 2014), an open source large-scale hydrological modeling system. It models the whole hydrological cycle: water balance, energy balance, snow melting, etc. (Figure 1). The system implements hydrological models, automatic calibration algorithms for model parameter optimization, and evaluation, and a GIS for input output visualization, (Formetta et al., 2011, Formetta et al., 2014). NewAge-JGrass is a component-based model: each hydrological process is described by a model (energy balance, evapotranspiration, run off production in figure 1); each model implement one or more component(s) (considering for example the model evapotranspiration in figure 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 section 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. Subsection 2.1 presents the landslide susceptibility model and subsection 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., 2015 are: the Montgomery and Dietrich (1994) model (M1), the Park et al. (2013) model (M2) and the Rosso et al., (2006) model (M3). The tree models derives from simplifications of the infinite slope
equation (Grahm J., 1984, Rosso et al., 2006, Formetta et al., 2014) for the factor of safety:

\[
FS = \frac{C \cdot (1 + e)}{\left(\frac{G_s + e \cdot S_r + w \cdot e \cdot (1 - S_r)}{G_s + e \cdot S_r + w \cdot e \cdot (1 - S_r)}\right)} \cdot \tan \phi' \cdot \tan \alpha
\]  

where \( FS \) is the factor of safety, \( C = C' + C_{\text{root}} \) is the sum of \( C_{\text{root}} \), the root strength [kN/m²] and \( C' \) the effective soil cohesion [kN/m²], \( \phi' [\cdot] \) is the internal soil friction angle, \( H \) is the soil depth [m], \( \alpha [\cdot] \) is the slope angle, \( \gamma_w \) [kN/m³] is the specific weight of water, and \( w = h/H \) \( [\cdot] \) where \( h \) [m] is the water table height above the failure surface [m], \( G_s \) [\cdot] is the specific gravity of soil, \( e [\cdot] \) is the average void ratio and \( S_r [\cdot] \) 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:

\[
w = \frac{h}{H} = \min\left(\frac{Q}{T} \cdot \frac{TCA}{b \cdot \sin \alpha \cdot 1.0}\right)
\]  

where \( T \) [L²/T] is the soil transmissivity defined as the product of the soil depth and the saturated hydraulic conductivity, \( b \) [L] is the length of the contour line.

Substituting eq. (2) in (1) the model is solved for \( Q/T \) assuming \( FS = 1 \) and stable and unstable sites are defined using threshold values on \( \log(Q/T) \) (Montgomery and Dietrich, 1994).

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

The component (M3) considers both the effects of rainfall intensity and duration on the landslide triggering process. The term \( w \) depends on rainfall duration and it is obtained by coupling the conservation of mass of soil water with the Darcy’s law (Rosso et al., 2006) providing:
Those models are suitable for shallow translational landslides controlled by groundwater flow convergence. Shallow landslides usually have a very low ratio between the maximum depth (D) and the length (L) of scar (D/L<0.1, Casadei et al., 2003), involve small volume of the colluvial soil mantle and present a generally translational failure mechanism (Milledge et al., 2014).

Each component has a user interface which specifies 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 (Abera 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 a 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 “unstable” in PL, but is “unstable” in OL, that is a missed alarm. The concept of the Receiver Operator

\[
W = \begin{cases} 
\frac{Q}{T} \frac{TCA}{b \cdot \sin \alpha} \left[ 1 - \exp \left( \frac{e + 1}{e} \left( \frac{TCA}{b \cdot \sin \alpha} \right)^{\ln \left( \frac{1 - T \cdot b \cdot \sin \alpha}{TCA \cdot Q} \right)} \right) \right] & \text{if } \frac{TCA}{b \cdot \sin \alpha} \leq \frac{e}{1 + e} \left( \frac{1 - T \cdot b \cdot \sin \alpha}{TCA \cdot Q} \right)^

\right]
\end{cases}
\]

\[W = \begin{cases} 
1 & \text{if } \frac{TCA}{b \cdot \sin \alpha} \geq \frac{e}{1 + e} \left( \frac{1 - T \cdot b \cdot \sin \alpha}{TCA \cdot Q} \right) \end{cases}
\]
Characteristic (ROC, Goodenough et al., 1974) graph is based on the values assumed by tp, fp, tn. The ROC is a methodology to assess the performance of models that provides results assigned to one of two classes. ROC graph is widely used in many scientific fields such as medicine (Goodenough et al., 1974), biometrics (Pepe, 2003) and machine learning (Provost and Fawcett, 2001). ROC graph is a Cartesian plane with the FPR on the x-axis and TPR on the y-axis. FPR is the ratio between false positive and the sum of false positive and true negative, and TPR is the ratio between true positive and the sum of true positive and false negative. They are defined in table 1 and commented in Appendix 1. The performance of a perfect model corresponds to the point P(0,1) on the ROC plane; points that fall on the bisector (black solid line, on the plots) are associated with models considered random: they predict stable or unstable cells with the same rate.

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

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

During the calibration procedure the selected algorithm compares model output in term of binary map (stable or unstable pixel) with the actual landslide optimizing a selected objective function (OF). The model parameter set for which the OF assumes its best value is the optimization procedure output. The eight GOF indices presented in table 1 were used in turn as OF and, consequently, eight optimal parameters sets were provided as calibration output (one for each optimised OF). To better clarify: a GOF index 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 assess what is (are) the OF index(es) whose optimization provides best model performances. Secondly, we verified if each OF metric has its 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. Subsection 3.1 describes the test-site; subsection 3.2 describes the model parameters calibration and verification procedure; subsection 3.3 presents the models performances correlations assessment; lastly, subsection 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 (Figure 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).

In the study area the topographic elevation has an average value of around 450 m a.s.l., with a maximum value of 730 m a.s.l. Slope, computed from 10 meters
The Crati Basin is a Pleistocene-Holocene extensional basin filled by clastic marine and fluvial deposits (Vezzani, 1968, Colella et al., 1987, Fabbricatore et al., 2014). The stratigraphic succession of the Crati Basin can be simply divided into two sedimentary units as suggested by Lanzafame and Tortorici, 1986. The first unit is a Lower Pliocene succession of conglomerates and sandstones passing upward into silty clays (Lanzafame and Tortorici, 1986) second unit. This is a succession of clayey deposits grading upward into sandstones and conglomerates referred to Emilian and Sicilian, respectively (Lanzafame and Tortorici, 1986), as also suggested by data provided by Young and Colella (1988). Mass movements were analyzed from 2006 to 2013 by integrating aerial photography interpretation acquired in 2006, 1:5000 scale topographic maps analysis, and extensive field survey. All the data were digitized and stored in GIS database (Conforti et al., 2014) and the result was the map of occurred landslide presented in figure 2,D. Digital elevation model, slope and total contributing area (TCA) maps are presented in figure 2, A, B, and C respectively. In order to perform model calibration and verification, the dataset of occurred landslides was divided in two parts one used for calibration (located in the bottom part of figure 2,D) and one for validation (located in the upper part of the figure 2,D). The landslide inventory map refers only to the initiation area of the landslides. This allows a fair comparison with the landslide models that provide only the triggering point and not include a runout model for landslides propagation.

3.2 Models calibration and verification

The three models presented in section 2 were applied to predict landslide susceptibility for the study area. Models’ parameters were optimized using each GOF index presented in table 1 in order to fit landslides of the calibration group. Table 2 presents the list of the parameters that will be optimized specifying their initial range of variation, and the parameters kept constant during the simulation and their value. The component PSO provides 8 best parameters set one for each optimized GOF indices. Values for each model (M1, M2 and M3) were presented in table 3. Optimal
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 m²/d) optimizing ACC is much lower compared to the transmissivity values obtained optimizing the other index (around 140 m²/d). Similar behavior is observed for the optimal rainfall value which is 148 [mm/d] optimizing ACC and around 70 [mm/d] optimizing the other indices. Considering the model M2, the optimal transmissivity and rainfall values optimizing CSI (10 [m²/d] and 95 [mm/d]), are much lower compared the values obtained optimizing the other indices (around 50 [m²/d] and 250 [mm/d] 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 models M1 and M2 could be due to compensate the effects of important physical processes neglected by those models.

Executing the models using the eight optimal parameters set, true-positive-rates and false positive rates are computed by comparing model output and actual landslides for both calibration and verification dataset. Results are 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 Appendix 2 both for calibration and for verification period. For the models M2 and M3 is clear that ACC, HSS, and CSI provide the less performing models results. This is true also for model M1, even if, differently form 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: FPR<0.4 and TPR>0.7. This choice was made in order to restrict the results’ comments only on the GOF indices that provide acceptable model results and for the readability of graphs. Figure 3 presents three ROC planes, one for each model, with the optimized GOF indices that provides FPR<0.4 and TPR>0.7. Results presented in Figure 3 and
Table 4 shows that: i) optimization of AI, D2PC, SI and TSS allows to reach the best model performance in the ROC plane, and this is verified for all three models; ii) performances increase as model complexity increases: moving from M1 to M3 points in the ROC plane approaches the perfect point (TPR=1, FPR=0); iii) increasing model complexity good model results are reached not only in calibration but also in validation dataset. In fact, moving from M1 to M2 soil cohesion and soil properties were considered, and moving from M2 to M3 rainfall of finite duration was used. The first step of the 3SVP procedure remarks that the optimization of AI, D2PC, SI, and TSS provides the best performances independently of the model we used.

3.3 Models performances correlations assessment

The second step of the procedure aims to verify the information content of each optimized OF, checking if it is analogous to other metrics or it is peculiar of the optimized OF. Executing a model using one of the eight parameters set (let’s assume, for example, the one obtained optimizing CSI) allows the computation of all the remaining GOF indices, that we indicate as CSI\(_{CSI}\), ACC\(_{CSI}\), HSS\(_{CSI}\), TSS\(_{CSI}\), AI\(_{CSI}\), SI\(_{CSI}\), D2PC\(_{CSI}\), ESI\(_{CSI}\), both for calibration and for verification dataset. Let’s denote this vector with the name \(MP_{CSI}\): the model performances (\(MP\)) vector computed using the parameters set that optimize CSI. \(MP_{CSI}\) has 16 elements, 8 for calibration and 8 for validation dataset. Repeating the same procedure for all eight GOF indices it gives: \(MP_{ACC}\), \(MP_{ESI}\), \(MP_{SI}\), \(MP_{D2PC}\), \(MP_{TSS}\), \(MP_{AI}\), \(MP_{HS}\). Figure 4 presents the correlation plots (Murdoch and Chow, 1996) between all \(MP\) vectors, for each model M1, M2 or M3. The matrix is symmetric and gives a certain ellipse at intersection of row i and column j. The color is the absolute value of the correlation coefficient between the \(MP_i\) and \(MP_j\) vectors. The ellipse’s eccentricity is scaled according to the correlation value: the more prominent the less the vectors 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 figure 4. This confirms that an optimization of AI, D2PC, SI and TSS provides 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 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.
- 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 were presented in Figures 5 and 6 for models M2 and M3 respectively. Each column of the figures represents one optimized index and has a number of boxplots 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 to a certain model parameters change. The narrower 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 from 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 other. Lastly, the third step shows 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 negative 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 flat response changing the parameters value. A more sensitive couple model-optimal parameter set will in fact accommodate eventual parameters, input data, or measured data variations responding to these changes with a variation of model performance. For this reason we used the combination the model M3 with parameters obtained by optimizing D2PC for drawing the final susceptibility maps in figure 7. Categories of landslides susceptibility from class 1 to 5 are assigned from low to high according to FS values (e.g. Huang et al., 2007): Class 1 (FS<1.0), Class 2 (1.0<FS<1.2), Class 3 (1.2<FS<1.5), Class 4 (1.5<FS<2.0), Class 5 (FS>2).

4 Conclusions

The paper presents a procedure to quantitatively calibrate, evaluate, and compare the performances of environmental models. The procedure was applied for the analysis of three landslides susceptibility models. It includes 3 steps: i) model parameters calibration optimizing different GOF indices and models evaluation in the ROC plane; ii) computation of degree of similarities between different models.
performances obtained by optimizing all the considered GOF index; iii) evaluation of models sensitivity to parameters variations.

The procedure has been conceived like a model configuration of the hydrological system NewAge-JGrass; it integrates: i) three simplified physically based landslides susceptibility models; ii) a package for model evaluations based on pixel-by-pixel comparison of modeled and actual landslides maps; iii) models parameters calibration algorithms, and iv) the integration with uDig open-source geographic information system for model input-output maps management.

This procedure was applied in a test case on the Salerno-Reggio Calabria highway and the best model performances were provided by model M3 optimizing D2PC index. In the application we presented the effective precipitation was calibrated because we were performing a landslide susceptibility analysis and it was useful for demonstrating the method. However, we are aware that for operational landslide early warning systems the rainfall constitutes a fundamental input of the predictive process. Moreover, the analysis would profit from measured rainfall data that triggered the occurred landslides, but that such data are not available at the moment for the study area.

The system is open-source and available at (https://github.com/formeppe). It is integrated according the Object Modeling System standards and this allows the user to easily integrate a generic landslide susceptibility model and use the complete framework presented in the paper avoiding rewriting programming code. The system could be improved by adding new landslide susceptibility models or different types of model selection procedure.

ACKNOWLEDGMENTS

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. The authors acknowledge the editor and the three reviewers (Prof. M. Mergili and two unknown reviewers) for providing insightful comments and improving the quality of the paper.
## Acronyms table

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

http://www.geomorphology.org.uk/sites/default/files/geom技術_chapters/2.4.1_GSToolbox.pdf


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


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


John Wiley & Sons.


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

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Range</th>
<th>Optimal value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical success index (CSI)</td>
<td>$CSI = \frac{tp}{tp + fn}$</td>
<td>[0, 1]</td>
<td>1.0</td>
</tr>
<tr>
<td>Equitable success index (ESI)</td>
<td>$ESI = \frac{tp - R}{tp + fn - R} = \frac{(tp + fn) - (tp + fn)}{(tp + fn + fp + fn + m)}$</td>
<td>[-1/3, 1]</td>
<td>1.0</td>
</tr>
<tr>
<td>Success Index (SI)</td>
<td>$SI = \frac{1}{2} \left( \frac{tp}{tp + fn} + \frac{fn}{fp + fn} \right)$</td>
<td>[0, 1]</td>
<td>1.0</td>
</tr>
<tr>
<td>Distance to perfect classification (D2PC)</td>
<td>$D2PC = \sqrt{(1 - TPR)^2 + FPR^2}$</td>
<td>[0, 1]</td>
<td>0.0</td>
</tr>
<tr>
<td>Average Index (AI)</td>
<td>$AI = \frac{1}{4} \left( \frac{tp}{tp + fn} + \frac{fp}{tp + fn} + \frac{fn}{fp + fn} + \frac{fn}{fn + m} \right)$</td>
<td>[0, 1]</td>
<td>1.0</td>
</tr>
<tr>
<td>True skill statistic (TSS)</td>
<td>$TSS = \frac{(tp - fn) - (fp - fn)}{(tp + fn) (fp + fn)}$</td>
<td>[-1, 1]</td>
<td>1.0</td>
</tr>
<tr>
<td>Heidke skill score (HSS)</td>
<td>$HSS = \frac{2 \cdot (tp - fn) - (fp - fn)}{(tp + fn) (fn + m) + (fp + fn) (fp + fn)}$</td>
<td>[-∞, 1]</td>
<td>1.0</td>
</tr>
<tr>
<td>Accuracy (ACC)</td>
<td>$ACC = \frac{tp + fn}{(tp + fn + fp + tn)}$</td>
<td>[0, 1]</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Table 2: Optimised models’ parameters values

<table>
<thead>
<tr>
<th>Model Parameters</th>
<th>Constant Value</th>
<th>Range value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil Depth [m]</td>
<td>-</td>
<td>[0.8; 5.0]</td>
</tr>
<tr>
<td>Transmissivity [m²/d]</td>
<td>-</td>
<td>[10; 150]</td>
</tr>
<tr>
<td>Soil/water density ratio</td>
<td>-</td>
<td>[1.8; 2.8]</td>
</tr>
<tr>
<td>Friction Angle [°]</td>
<td>-</td>
<td>[11; 40]</td>
</tr>
<tr>
<td>Rainfall [mm/d]</td>
<td>-</td>
<td>[50; 300]</td>
</tr>
<tr>
<td>Soil Cohesion [kPa]</td>
<td>-</td>
<td>[0; 50]</td>
</tr>
<tr>
<td>Degree Of Saturation [-]</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td>Soil Porosity [-]</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td>Rainfall Duration [d]</td>
<td>-</td>
<td>[0.1; 3.0]</td>
</tr>
</tbody>
</table>
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).

<table>
<thead>
<tr>
<th>Optimised Index</th>
<th>AI</th>
<th>HSS</th>
<th>TSS</th>
<th>D2PC</th>
<th>SI</th>
<th>ESI</th>
<th>CSI</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil Depth [m]</td>
<td>1.32</td>
<td>1.85</td>
<td>1.44</td>
<td>2.80</td>
<td>1.36</td>
<td>2.62</td>
<td>2.42</td>
<td>2.01</td>
</tr>
<tr>
<td>Transmissivity [m²/d]</td>
<td>140.24</td>
<td>146.31</td>
<td>142.68</td>
<td>137.10</td>
<td>147.69</td>
<td>144.66</td>
<td>136.73</td>
<td>74.74</td>
</tr>
<tr>
<td>Soil/water density ratio [-]</td>
<td>2.61</td>
<td>2.56</td>
<td>2.77</td>
<td>2.71</td>
<td>2.78</td>
<td>2.79</td>
<td>2.63</td>
<td>2.72</td>
</tr>
<tr>
<td>Friction Angle [°]</td>
<td>24.20</td>
<td>32.40</td>
<td>22.50</td>
<td>23.10</td>
<td>22.40</td>
<td>29.50</td>
<td>29.50</td>
<td>74.74</td>
</tr>
<tr>
<td>Rainfall [mm/d]</td>
<td>85.38</td>
<td>53.30</td>
<td>71.36</td>
<td>50.00</td>
<td>62.69</td>
<td>69.19</td>
<td>61.35</td>
<td>141.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Optimised Index</th>
<th>AI</th>
<th>HSS</th>
<th>TSS</th>
<th>D2PC</th>
<th>SI</th>
<th>ESI</th>
<th>CSI</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmissivity [m²/d]</td>
<td>65.43</td>
<td>33.22</td>
<td>80.45</td>
<td>38.22</td>
<td>84.54</td>
<td>33.24</td>
<td>10.70</td>
<td>55.76</td>
</tr>
<tr>
<td>Cohesion [kPa]</td>
<td>25.17</td>
<td>49.63</td>
<td>49.42</td>
<td>16.94</td>
<td>30.01</td>
<td>41.24</td>
<td>44.58</td>
<td>46.85</td>
</tr>
<tr>
<td>Friction Angle [°]</td>
<td>29.51</td>
<td>38.38</td>
<td>20.01</td>
<td>32.30</td>
<td>24.57</td>
<td>33.78</td>
<td>35.68</td>
<td>34.96</td>
</tr>
<tr>
<td>Rainfall [mm/d]</td>
<td>236.14</td>
<td>293.44</td>
<td>270.42</td>
<td>153.61</td>
<td>294.70</td>
<td>298.44</td>
<td>95.35</td>
<td>299.01</td>
</tr>
<tr>
<td>Soil/water density ratio [-]</td>
<td>2.11</td>
<td>2.40</td>
<td>2.06</td>
<td>2.44</td>
<td>2.77</td>
<td>2.17</td>
<td>2.55</td>
<td>2.19</td>
</tr>
<tr>
<td>Soil Depth [m]</td>
<td>2.35</td>
<td>1.68</td>
<td>2.38</td>
<td>2.44</td>
<td>2.74</td>
<td>1.12</td>
<td>1.37</td>
<td>1.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Optimised Index</th>
<th>AI</th>
<th>HSS</th>
<th>TSS</th>
<th>D2PC</th>
<th>SI</th>
<th>ESI</th>
<th>CSI</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmissivity [m²/d]</td>
<td>30.95</td>
<td>26.55</td>
<td>47.03</td>
<td>36.31</td>
<td>57.28</td>
<td>25.84</td>
<td>31.60</td>
<td>48.71</td>
</tr>
<tr>
<td>Cohesion [kPa]</td>
<td>36.88</td>
<td>44.33</td>
<td>28.51</td>
<td>31.60</td>
<td>45.46</td>
<td>41.80</td>
<td>32.05</td>
<td>37.09</td>
</tr>
<tr>
<td>Friction Angle [°]</td>
<td>19.55</td>
<td>36.44</td>
<td>27.80</td>
<td>29.70</td>
<td>21.46</td>
<td>33.27</td>
<td>36.47</td>
<td>38.50</td>
</tr>
<tr>
<td>Rainfall [mm/d]</td>
<td>248.77</td>
<td>230.08</td>
<td>258.82</td>
<td>201.71</td>
<td>299.90</td>
<td>291.32</td>
<td>273.03</td>
<td>193.02</td>
</tr>
<tr>
<td>Soil/water density ratio [-]</td>
<td>2.40</td>
<td>2.57</td>
<td>2.08</td>
<td>2.80</td>
<td>2.65</td>
<td>2.63</td>
<td>2.61</td>
<td>2.44</td>
</tr>
<tr>
<td>Soil Depth [m]</td>
<td>1.84</td>
<td>1.42</td>
<td>2.23</td>
<td>2.92</td>
<td>2.85</td>
<td>1.17</td>
<td>1.13</td>
<td>1.15</td>
</tr>
<tr>
<td>Rainfall Duration [d]</td>
<td>0.12</td>
<td>1.78</td>
<td>1.24</td>
<td>1.96</td>
<td>1.24</td>
<td>0.39</td>
<td>1.30</td>
<td>1.98</td>
</tr>
</tbody>
</table>
### Table 4: Results in term of true-positive rate (TPR) and false-positive rate (FPR), for each model (M1, M2 and M3), for each optimised GOF index and for both calibration (CAL) and verification (VAL) dataset. In bold are shown the rows for which the condition FPR<0.4 and TPR>0.7 is verified.

<table>
<thead>
<tr>
<th>Period</th>
<th>Optim. Index</th>
<th>MODEL: M1</th>
<th>MODEL: M2</th>
<th>MODEL: M3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FPR TPR</td>
<td>FPR TPR</td>
<td>FPR TPR</td>
<td>FPR TPR</td>
</tr>
<tr>
<td>CAL</td>
<td>ACC</td>
<td>0.04 0.12</td>
<td>0.03 0.12</td>
<td>0.03 0.13</td>
</tr>
<tr>
<td>CAL</td>
<td>AI</td>
<td>0.29 0.70</td>
<td>0.35 0.79</td>
<td>0.38 0.82</td>
</tr>
<tr>
<td>CAL</td>
<td>CSI</td>
<td>0.17 0.48</td>
<td>0.10 0.36</td>
<td>0.09 0.32</td>
</tr>
<tr>
<td>CAL</td>
<td>D2PC</td>
<td>0.32 0.72</td>
<td>0.32 0.76</td>
<td>0.32 0.75</td>
</tr>
<tr>
<td>CAL</td>
<td>ESI</td>
<td>0.17 0.48</td>
<td>0.43 0.82</td>
<td>0.09 0.36</td>
</tr>
<tr>
<td>CAL</td>
<td>HSS</td>
<td>0.12 0.35</td>
<td>0.09 0.35</td>
<td>0.09 0.35</td>
</tr>
<tr>
<td>CAL</td>
<td>SI</td>
<td>0.34 0.74</td>
<td>0.39 0.85</td>
<td>0.39 0.86</td>
</tr>
<tr>
<td>CAL</td>
<td>TSS</td>
<td>0.34 0.73</td>
<td>0.39 0.83</td>
<td>0.37 0.82</td>
</tr>
<tr>
<td>VAL</td>
<td>ACC</td>
<td>0.05 0.12</td>
<td>0.03 0.12</td>
<td>0.03 0.10</td>
</tr>
<tr>
<td>VAL</td>
<td>AI</td>
<td>0.26 0.56</td>
<td>0.31 0.69</td>
<td>0.34 0.72</td>
</tr>
<tr>
<td>VAL</td>
<td>CSI</td>
<td>0.17 0.39</td>
<td>0.09 0.31</td>
<td>0.08 0.29</td>
</tr>
<tr>
<td>VAL</td>
<td>D2PC</td>
<td>0.29 0.59</td>
<td>0.28 0.67</td>
<td>0.28 0.66</td>
</tr>
<tr>
<td>VAL</td>
<td>ESI</td>
<td>0.17 0.39</td>
<td>0.41 0.76</td>
<td>0.09 0.30</td>
</tr>
<tr>
<td>VAL</td>
<td>HSS</td>
<td>0.12 0.30</td>
<td>0.09 0.30</td>
<td>0.09 0.30</td>
</tr>
<tr>
<td>VAL</td>
<td>SI</td>
<td>0.30 0.61</td>
<td>0.37 0.75</td>
<td>0.39 0.76</td>
</tr>
<tr>
<td>VAL</td>
<td>TSS</td>
<td>0.30 0.62</td>
<td>0.35 0.74</td>
<td>0.34 0.71</td>
</tr>
</tbody>
</table>
**Figure 1:** Integration of the Landslide susceptibility analysis system in NweAge-JGrass hydrological model.
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.
**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 range 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.

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

\[
R = \frac{(tp + fn)(tp + fp)}{tp + fn + fp + fn} \quad (4)
\]

1.4 Success index (SI)

SI, eq. (5), equally weight True positive rate (eq. 6) and specificity defined as 1 minus false positive rate (FPR), eq. (7). SI varies between 0 and 1 and its best value is 1. SI is also named modified success rate.
\[ \text{SI} = \frac{1}{2} \left( \frac{\text{tp}}{\text{tp} + \text{fn}} + \frac{\text{tn}}{\text{fp} + \text{fn}} \right) = \frac{1}{2} \times (\text{TPR} + \text{specificity}) \] (5)

\[ \text{TPR} = \frac{\text{tp}}{\text{tp} + \text{fn}} \] (6)

\[ \text{FPR} = \frac{\text{fp}}{\text{fp} + \text{tn}} \] (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.

\[ D2PC = \sqrt{(1 - \text{TPR})^2 + \text{FPR}^2} \] (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).

\[ AI = \frac{1}{4} \left( \frac{\text{tp}}{\text{tp} + \text{fn}} + \frac{\text{tp}}{\text{tp} + \text{fp}} + \frac{\text{tn}}{\text{fp} + \text{tn}} + \frac{\text{tn}}{\text{fn} + \text{tn}} \right) \] (9)

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

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

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

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

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

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

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

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

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

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 tn is large the false alarm value is relatively overwhelmed. If tn is large, as happens in landslides maps, FPR tends to zero and TSS tends to TPR. A problem of TSS is that it treats the hit rate and the false alarm rate equally, irrespective of their likely differing consequences.

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

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.