Reply to reviewer n.3: M. Mergili

"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 eaerly 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

Bischetti, G. B., & Chiaradia, E. A. (2010). Calibration of distributed shallow landslide models in forested landscapes. *Journal of Agricultural Engineering*, *41*(3), 23-35.

Deb, S. K., & El-Kadi, A. I. (2009). Susceptibility assessment of shallow landslides on Oahu, Hawaii, under extreme-rainfall events. Geomorphology, 108(3), 219-233.

Huang, J. C., and S. J. Kao. "Optimal estimator for assessing landslide model performance." *Hydrology and Earth System Sciences Discussions* 10, no. 6 (2006): 957-965.