

Nat. Hazards Earth Syst. Sci. Discuss., author comment AC2
<https://doi.org/10.5194/nhess-2021-140-AC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

Feiko Bernard van Zadelhoff et al.

Author comment on "Introducing SlideforMAP: a probabilistic finite slope approach for modelling shallow-landslide probability in forested situations" by Feiko Bernard van Zadelhoff et al., Nat. Hazards Earth Syst. Sci. Discuss.,
<https://doi.org/10.5194/nhess-2021-140-AC2>, 2021

We would like to thank the reviewer for the overall positive assessment of our paper and the detailed suggestions to further improve the manuscript. We give in italic below the original comment and in normal font our answers below. We did not copy the entire general comment part from the review, which provided a nice summary of our work (thanks for this!) but did not require any specific answer.

GENERAL COMMENTS

The authors describe a probabilistic model called SlideforMap (SfM) which generates a map of shallow landslide probability across an area of interest. (..) The authors have organized their manuscript well, and they have described a complex workflow in a straightforward way. They also build a convincing case for the utility and need for a model of this type, and the described case studies illustrate the applications well. In my opinion, this manuscript should be published in NHSS after some clarifications and revisions. Most of my criticisms are focused on areas where the authors need to provide additional clarifications, either to adequately explain their approach or to explain how this model could be used by others.

Thanks for the nice summary and the positive assessment. The revised paper will contain more details on how the model could be used by others (see also further comments below and response to RC1, 6. Results question).

SPECIFIC COMMENTS

1) *The authors say that their model demonstrates the importance of root reinforcement on*

shallow landslides, but the authors need to define what "shallow" means so that it is clear where their conclusions apply.

An informal definition of shallow landslides is landslides within the soil mantle, not containing bedrock. This is different from the official swiss definition, which states these are landslides with a soil depth < 2 m. In this paper, we officially use the Swiss definition, however, for the practical application of SlideforMap we consider this definition to be irrelevant. We will state this more clearly in the introduction and conclusion.

The authors are persuasive about the importance of root reinforcement in modeling landslide hazards, but they do not provide much discussion of how this model compares to other previously published models, including both related models (such as SOSlope or SlideForNET) or other models that compute landslide susceptibility on a regional scale. Some additional discussion of where this model fits within the context of other landslide susceptibility models generally would be helpful for prospective users.

Thank you for this point. The authors agree this issue can be better emphasized in the discussion. We will add a paragraph explicitly comparing SlideforMap to other landslide susceptibility models in their application.

In describing the methodology, the authors are not always clear about which values are assumed for their own case study, and which values are fixed in the model. For example, at a number of places in the methodology section, the authors assign values and limits on parameters (e.g., maximum HL surface area, mean tree density, precipitation intensity threshold, etc.) based on data from Switzerland (where the case study is located), but it is not clear whether a given user would have the freedom to change these values.

Thank you for pointing this out with the concrete examples. The thresholds are indeed Switzerland specific, others are based on assumptions by the authors. Future users have the opportunity to change these values and are encouraged to do so if they apply SlideforMap in other areas. We will make clear in the revised version which parameters are specifically selected for Switzerland and which ones are related to more general assumptions.

The structure of the model requires that soil depth, soil cohesion, and the angle of internal friction be modeled as random variables with normal distributions, but the other 16 parameters are assumed to be deterministic. The authors need to explain why these three parameters specifically were chosen to be random variables. For instance, variables can be randomized when the uncertainty in their values is either shown or assumed to have the most significant effects on the results. This is suggested somewhat by the sensitivity analysis for the case of soil cohesion and soil depth, but this choice is not explained explicitly.

Indeed as the reviewer suggests this was partly motivated by sensitivity. Soil depth and soil cohesion are generally assumed to be highly influential in shallow landslide susceptibility mapping (e.g. Cislighi et al., 2018).

We will make clearer in the revised version the difference between i) applying a random parameter field for selected parameters (i.e. different randomly generated parameter values for each grid cell) and ii) generating a set of random parameters for parameter calibration. These are two very different, unrelated stages of parameter identification, where the first one is a way of assigning spatially heterogeneous parameters values and the second a way of identifying good or best parameter estimates.

The motivation for applying random fields is the application in hilly or mountainous areas for SlideforMap. Here, soil properties are generally highly heterogeneous (e.g. Tofani et al., 2017) and we would like to account for this specific heterogeneity in the probabilistic approach of SlideforMap. Two soil properties are left out intentionally. Firstly soil density, which is assumed to have low spatial variability and influence. Secondly soil transmissivity, which is considered part of the hydrological approach and is included in calibration. We will explain this choice more explicitly in the revised version.

The authors make use of two datasets, a tree inventory and a landslide inventory, in their analysis. However, they do not spend much time explaining how a prospective user would apply this model if they were lacking these datasets. It seems that users could still apply this model without these datasets, either by creating synthetic datasets or assuming specific values for the parameters that would be derived from these datasets. Providing some more guidance on applying the model without these datasets this would make the model more accessible to users.

Good point. Both synthetic datasets and assuming specific values (e.g. distribution parameters taken directly from the Malamud et al., (2004) are possible. We will make this clearer in the SlideforMap section (section 2) of the paper.

The sensitivity analysis is interesting but not entirely convincing. If strong parameter correlation is at play, as the authors suggest, then how would we know which parameters are truly important?

We did not want to imply that strong parameter correlation is at play. Figure 1 in the attachment shows pair-wise dotted plots between all parameters. Figure 2 below shows all the pair-wise linear correlation coefficients. Both figures are for the 20% best parameter sets. These figures show a correlation between mean depth - mean cohesion, Transmissivity- mean cohesion and Friction angle - mean cohesion. This indicates that the influence of these parameters on the performance may be higher than comes forward in our manuscript figure 7. In addition, further multi-variate correlations or bivariate non-linear correlations can be at play as well.

Attachment Figure 1: pair-wise dotted plots between the 20 % best parameters according to AUC

Attachment Figure 2: pair-wise correlation between the 20 % best parameters according to AUC

What we intended to say in the original manuscript is that a *potential* correlation between parameters can lead to apparent absence of sensitivity. This was nicely demonstrated in the paper by Bardossy (2007). We reproduce below that example (see full Matlab code as a supplement to this response) where we try to find the best parameters of a simple Nash cascade model fitted to a hydrograph that we generated ourselves with that model. The performance criteria (here the RMSE) does not show a strong sensitivity with respect to the randomly sampled parameters (Figure 3). This result is due to the correlation between the best parameter sets (Attachment figure 3 right). What this example shows is that if parameter sets show a correlation this *might* manifest itself in absence of sensitivity. Accordingly, absence of sensitivity does not imply that the parameters have to be correlated.

Attachment Figure 3: left and center: scatter plot of parameter values against performance measure of the synthetic Nash cascade example, showing an apparent low sensitivity; right: plot of the best 20% of all parameters sets against each other showing the strong correlation between the best parameter sets.

In a couple of places within the text (L49-52; L169-170) the authors conflate deterministic models with spatial homogeneity. This is misleading, as it is possible to have deterministic models that account for spatial heterogeneity, and probabilistic models that are spatially homogeneous. I would suggest that the explanation the authors are after is that the spatially heterogeneous values themselves are uncertain, and this is the motivation for using a probabilistic approach.

This is a good point. We will adjust the text.

Is it valid to compare the globally uniform vegetation scenario to the other three scenarios if the globally uniform scenario was used to calibrate the parameters?

Good point. Indeed we would argue this gives an 'unfair advantage' to the uniform vegetation scenario. Performance is comparable (in the case of the Trub study area even better) with the single tree detection detection. In our opinion this strengthens our case that the model can be calibrated on uniform vegetation and then be used with different vegetation scenarios.

It appears that the authors used the same landslide inventory to both calibrate the dataset and to validate the performance of the model against different vegetation scenarios. Did the authors consider using any portion of the landslide inventory as an independent validation dataset?

We wanted to analyze the performance of the model, not do a validation. For a proper validation we think the size of the dataset is too limited. The vegetation scenarios are used to demonstrate the usefulness of the model to assess the impact different vegetation scenarios but not to validate the model (in the context of having a set of best performing parameters). This will be made clearer in the revised version.

L44-45. The authors need to give some additional definition of a deterministic approach and why SHALSTAB is an example of this approach.

Good point and related to the comment on spatial homo/heterogeneity. In the same section we will more precisely define deterministic and probabilistic approaches and what it does and does not entail.

L128-130. It seems that the unstable ratio is a very limited metric, particularly if the landslide density is already very low. Shouldn't the landslide density be relevant in addition to the unstable ratio? If there is an explicit requirement that the number of HLs be large enough to compute the unstable ratio with a large denominator, does this effectively put a lower bound on the landslide density for this model?

Both the unstable ratio as a metric and the AUC are influenced by the landslide density. It can also be seen as a slight factor of influence in the sensitivity analysis (Figure 7). Although the landslide density definitely has a lower bound for reliability, we have not specifically identified this boundary. We considered this out of the scope of this research. We agree that the unstable ratio in general is a limited metric since it is not an independent metric of performance, therefore we choose the AUC as main metric.

L152-153. I am surprised that the landslides are generated using a spatially uniform distribution, as this may result in landslides being simulated in areas that are not landslide prone. What is the rationale behind this? Shouldn't they follow a spatially distributed density, or at least be restricted to susceptible areas?

This was chosen to be unbiased in the definition of what was qualified as susceptible. This choice however influences the AUC metric as we discuss in section 5.5. This is a constraint many natural hazard model studies struggle with (Corominas et al., 2014), but to stay in line with similar model publications we decided to keep it as our performance metric.

L278. A 2km buffer seems extremely large, especially if topographic wetness is computed over multiple small catchments. How was this value chosen, and is it adequate for other studies?

The value was deliberately chosen large to avoid any doubt in the TWI accuracy. As noticed it is extremely large, but we had the DEM available to do it and the GIS procedure is not complicated. Model users are free to use a smaller buffer if they are confident it will still result in a correct TWI computation. This will be specified in the revised version.

L407-408. What does this mean if the unstable ratio decreases when single tree detection is used? Does this indicate that heterogeneity is important for slope stability, or does it simply mean that the uniform vegetation scenarios are not realistic?

The most realistic explanation is that root reinforcement from single tree detection exceeds that of the calibrated uniform vegetation scenario. When applied in susceptible areas, it decreases instability to a greater extent than it would in the uniform vegetation scenario. It is an issue of both placement and amount. This will be added to the discussion.

Table 7. Why are the AUC and Unstable ratio values different for the globally uniform vegetation scenario compared to the results with the optimal parameters (Table 6)? Is this due to the difference in the landslide density?

Table 7 takes the average of 10 runs with identical model parameters but different realisations with random placement of the landslides. We will make this clearer. Table 6 reports the result corresponding to a single run with the best parameter set of the 1000 randomly generated sets (for each of which a single run was computed).

L471-473. Does this high unstable ratio match with long term observations about landslide occurrence in StA? In other words, is the unstable ratio realistic?

This is a good question and can only to a certain extent be analyzed from our landslide inventory. From the inventoried slides and the surface area of the study areas a landslide density can be computed. These would be in slides/km²: Eriz: 4.9, Trub; 8, StA: 58.9. These results indicate a higher landslide density in the StA area comparable to our results. Insecurity arises of course due to the fact that events can not be 1:1 compared, but it could give a rough estimate. We can add this column in the table and shortly mention this in the discussion.

L480. This suggests that AUC is a poor choice of performance metric for comparing the three study areas. Are there other metrics which would be better?

That is a good point. As stated, though the AUC is frequently used, it has its shortcomings. This problem has been analyzed and better (though not optimal) propositions have been made (e.g. Chung & Fabbri, 2003). However in order to compare the results in an easy manner to performance of other models, we decided to stick with the AUC. In a future paper, we would like to diversify the performance evaluation of SfM, but we consider this outside the scope of presenting the model in the first place.

TECHNICAL CORRECTIONS

Reply: thanks for pointing out the corrections, we will implement them.

LITERATURE

Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences*, 11(2), 703–710.
<https://doi.org/10.5194/hess-11-703-2007>

Chung, C. J. F., & Fabbri, A. G. (2003). Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*, 30(3), 451–472.
<https://doi.org/10.1023/B:NHAZ.0000007172.62651.2b>

Cislaghi, A., Rigon, E., Lenzi, M. A., & Bischetti, G. B. (2018). A probabilistic multidimensional approach to quantify large wood recruitment from hillslopes in mountainous-forested catchments. *Geomorphology*, 306, 108–127.
<https://doi.org/https://doi.org/10.1016/j.geomorph.2018.01.009>

Corominas, J., vanWesten, C., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi,

F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J., and Smith, J. T.: Recommendations for the quantitative analysis

of landslide risk, *Bulletin of Engineering Geology and the Environment*, 73, 209–263,
<https://doi.org/10.1007/s10064-013-0538-8>, 2014.

Malamud, B., Turcotte, D., Guzzetti, F., & Reichenbach, P. (2004). Landslide inventories and their statistical properties. *Earth Surface Processes and Landforms*, 29(6), 687–711. <https://doi.org/10.1002/esp.1064>

Tofani, V., Bicchocchi, G., Rossi, G., Segoni, S., D'Ambrosio, M., Casagli, N., & Catani, F. (2017). Soil characterization for shallow landslides modeling: a case study in the Northern Apennines (Central Italy). *Landslides*, 14(2), 755–770. <https://doi.org/10.1007/s10346-017-0809-8>

Please also note the supplement to this comment:

<https://nhess.copernicus.org/preprints/nhess-2021-140/nhess-2021-140-AC2-supplement.zip>