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 approach is to randomly simulate “hypothetical landslides” across the landscape. A factor of safety is calculated for each hypothetical landslide based on limit-equilibrium analysis. Areas with factors of safety less than 1 are assumed to be unstable. The final output of the model is the fraction of unstable landslides at a given location.

The model requires a total of 22 parameters, 16 of which are deterministic and 3 of which are drawn from probability distributions (each of which is defined by 2 parameters). The resisting force is based partially on pore water pressures computed using a variation of TOPMODEL, which is justified by assuming the dominant role of macropore flow in pore pressure development.

The novel feature promoted by the authors is the inclusion of basal and lateral root reinforcement from vegetation into the resisting force. In their case study, the authors demonstrate how they obtain an inventory of trees from an airborne laser scanning dataset. The authors also argue that the root reinforcement should explicitly depend on the size and spatial distribution of individual trees.

The authors demonstrate their approach using three study areas in Switzerland, each of which has a landslide inventory that the authors use to calibrate the model against. The authors conduct a sensitivity analysis where they analyze how the marginal distributions of different parameters are related to model performance and to the fraction of unstable landslides.

The authors include a thorough discussion that explores the limitations of the reference datasets in assessing model performance, the assumptions of several model parameterization choices, and the AUC as a performance metric. Much of this discussion reflects on the differences in the predicted unstable ratio between one of the study areas, St Antonien, and the other two.

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 NHESS 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.

**SPECIFIC COMMENTS**

- 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.
- 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.
- 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.
- 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.
- 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 would make the model more accessible to users.
- 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?
- 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.
- 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?

- 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?

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

- 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?

- 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?

- 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?

- 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?

- 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?

- 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?

- 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?

TECHNICAL CORRECTIONS

- L14. This should be “ratio” instead of “fraction.”

- L121. Does SfM generate a raster image of probability values?

- L134. Do the authors mean “greater than 1.0”?

- L163. What does “distance of 10” refer to?

- L 271-273. What resolution is the unstable ratio computed at? This is not made explicit here in the paper.

- L300-307. What is the spatial format of the landslide inventory? If they are polygons, how are they compared to the unstable ratio map so that the AUC can be computed? Does the landslide inventory need to be converted or rasterized at a specific resolution?

- L309. The format for the numbers a, b, and c looks unusual. Please verify that the values and formats are correct.

- L312-313. Why are these 11 parameters fixed while the others are varied?

- L332. What is n?
L336-337. Please explain why weighting is being used and how this weighting is calculated.
L343. Please explain why the parameter range is using intensity values from different return periods.
Table 5. The value for vegetation weight, Wveg, uses a different name and different units than the rhotree in Table 1 (tonne per square meter vs. kg per cubic meter). Is there a reason for this difference?
L348. Is 1000 an adequate size to represent the sample space over the 12 parameters used in the sensitivity analysis?
L360-361. Does this model assume that root reinforcement comes only from trees, and not from shrubs, grasses, or other vegetation types? Is the single-tree detection scenario using the same trees as the tree inventory cited in 3.2?
L363. Please verify that the exponent is correct in the expression for landslide density.
Fig. 8. How is "x% best" defined for the unstable ratio?
L403. Do the model runs assume randomization of the three parameters (as in the original model setup)?
L508. Are the 12 parameters all included in the 22 original parameters?