We would like to thank CC1 for taking the time review our manuscript, the provided feedback and the clear suggestions for improvement. In italic below we give the original comment and in normal font our answers.

**the model itself is similar to a number of existing models but also makes some important changes. It would be really useful to make these similarities and differences more explicit.**

The striking similarities to me were: 1) the hydrological model (Eqns 11-12) is exactly that of SHALSTAB (Montgomery and Dietrich, 1994) and SINMAP (Pack et al., 1998);  2) modelling discrete landslides of defined dimensions with lateral resistance due to roots only (Eqns 1-6) follows Montgomery et al. (2000), Schmidt et al. (2001) and Roering et al. (2003); 3) the probabilistic treatment of stability using distributions for parameters follows Pack et al. (1998) who represented c, phi and the R/T ratio as uniform distributions; 4) introducing a slope dependence to failure depth follows Prancevic et al. (2020), though with a different functional form. The similarities are strongest between SfM and Montgomery et al. (1998), they use very similar stability models (both infinite slope with root cohesion only on the margins), the same hydrological model, and both impose discrete landslide dimensions; so differentiating your work from theirs will be important.

Thank you for pointing out the similarities in approach of SlideforMap in this systematic way.

1) you’re right. We use the same hydrological approach as SHALSTAB. We will specify this better in the paper.

2) In our opinion, the concept is identical, but the implementation is different. Our approach to the quantification of root reinforcement and the probabilistic approach for
landslide dimensions (eq. 6) is a unique integration. Our lateral root reinforcement is spatially heterogeneous and multiplied by half of the landslide perimeter instead of the whole perimeter. This is closer to reality according to Schwarz et al., (2010) and in contrast to Montgomery et al., (2000), Roering et al., (2003) and Schmidt et al., (2001) who use the full perimeter of the landslide. We employ a probabilistic approach to the landslide surface areas by sampling from a distribution. This is in contrast to the fixed width (5m) and length (10m) of the landslides in Montgomery et al., (2000). In addition, in contrast to the papers commented, we employ the distinct Root Bundle Model (RBMw) (Gehring et al., 2019) to estimate spatial lateral and basal root reinforcement.

3) Yes that is correct, even though we use the normal distribution for soil parameters as opposed to the uniform distribution in Pack et al., (1998).

4) The approach of fitting landslide scar depth to slope angle is indeed similar to Prancevic et al. (2020). In Prancevic et al., (2020) it is used to define relationships and dimensions for slope angle domains, we use it as a correction factor to keep our probabilistic samples realistic.

Although SfM is similar in many aspects of the approaches, the main difference is related to the quantification of root reinforcement (Our spatial distributed root reinforcement comes specifically from forest structure) and in the implementation of the probabilistic computation. Compared to Montgomery et al., (1998), the hydrological module is identical. We choose the application of this model as a result of observation in an artificial precipitation experiment by Askarinejad et al., (2012). But in other modules of the model we have notable differences. We will add specific similarities and distinction to our method section in the revised paper and discuss these in the discussion section.

Having read the paper I have one primary outstanding question: What do you gain as a result of the additional data collection and modelling efforts involved in a detailed inclusion of the influence of vegetation?

The opportunity to work with single tree detection based vegetation parameters opens up opportunities for users to assess the impact of different vegetation scenarios. This will have advantages since forest management and/or reforestation are an important if not the most important mitigation strategy there is for shallow landslides in many areas (e.g. Amishev et al., 2014). Moreover, we expect an improvement in the performance of the model with the detailed inclusion of vegetation, which we analyzed in the paper.

Your paper focuses on predictive skill (using ROC AUC) and predicted instability (using an unstable area ratio). That focus enables a straightforward assessment of improvement in predictive skill from this more complex model relative to a simpler models such as SHALSTAB or SINMAP. In fact, I think you already have an answer to this in Table 7. The ‘no vegetation’ case in SIM is very close to the SINMAP model: in this case, there is no lateral resistance (i.e. an infinite slope), probability of failure is calculated from pdfs of friction, cohesion and depth with pore pressure predicted using the SINMAP/SHALSTAB model. The uniform vegetation cases (Global and Forest area) are very close to the SHALSTAB implementation of Montgomery et al. (2000): in these cases landslides have predefined dimensions and lateral cohesion is spatially uniform. The difference is that
landslide dimensions (area and depth), and material properties (c and phi) are sampled from distributions to generate a probability of failure rather than using the critical P/T as a metric for propensity to failure (as in SHALSTAB). In all these cases I would expect a direct comparison to SINMAP and the SHALSTAB of Montgomery et al. (2000) to yield almost exactly the same AUCs as those from SfM. The clear structural difference between SfM and previous models comes in the case of ‘Single tree detection’.

Thank you for pointing this out, adding this explicitly will be a good addition to the paper. However, we would like to note that there are further differences to the model approach as pointed out in the previous questions. In addition we integrate the spatial parameters (slope, topographic wetness index, lateral root reinforcement) over all cells in the landslide surface area. The landslide surface area samples are drawn from a calibrated surface area distribution.

Reading Table 7 in the context of these connections to simpler early models leads to three conclusions:

- Landslide predictions are surprisingly (and encouragingly) skilful even when models as simple as the ‘No vegetation’ SfM (equivalent to SINMAP) are used. Models like SINMAP are very attractive if they perform so well given their simple structure and parsimonious parameterisation.

- Representing landslides as discrete features (as in SfM or Montgomery et al. (2000)) rarely improves predictive skill unless detailed vegetation information is available. Best AUC for SfM with ‘Global’ or ‘Forest area vegetation’ are equal to the ‘No vegetation’ case for 2 of the 3 study sites and only 1% better for Sta.

- Detailed vegetation information from single tree detection does subtly improve predictive skill but only in 2 of the 3 sites (slightly worse for Eriz) and only by 3.8 and 3.2% in AUC for Trub and Sta respectively.

One interpretation of this would be that while SfM is much more satisfying from a process representation point of view it offers only very marginal gains in predictive skill and has considerable cost in that it is more highly parameterised and more complex. An alternative interpretation would be that small skill improvements on an already excellent model are worth the additional complexity (and cost). Reframing the percentage changes in AUC as percentage of the unrealised AUC that has been eroded by the new model (thus changing in denominator from $AUC_{\text{pre}}$ to $1-AUC_{\text{pre}}$) the same values are: 6% and 43% for Trub and Sta respectively. I think this interpretation, which recognises the diminishing returns in model improvement is reasonable and if so it suggests the improvement is non-trivial.

It is interesting that the unstable ratio metric is more sensitive to model structure than AUC, and perhaps encouraging that this ratio is reduced by improved process representation. However as you point out (L355), this ratio is a measure of instability rather than accuracy.
The authors agree to a large degree with this interpretation and would like to thank CC for this interesting and concise discussion. We will add this discussion to the paper and acknowledge CC explicitly for the contribution to the improvement of this paper.

With single tree detection, the relative gain of the AUC is marginal, but as pointed out in the comment, the decrease in unstable ratio is more significant. This is noteworthy and we will add some possible explanations in the discussion, as we are not certain of the causes. It can possibly be further reduced by running future scenario’s with denser, well managed forest or reforestation efforts in areas where there is none as of yet. This is also the main ability of SlideforMap, to quantify how instability can be reduced under different forest scenarios.

Ps. CC1 swapped study areas in the comment: Sta has 6% percentage change in AUC, Trub has 43%.

**SfM also makes predictions about the size of landslides** most likely to be triggered in each location (though these are not currently reported in the paper). This is an important difference from previous models. Few models have done this before and those that have are extremely computationally expensive. Therefore the most exciting aspect of SfM to me is its ability to predict landslide size. The authors are clear that the model requires a prior distribution of landslide sizes but this does not prevent SfM from producing useful information on landslide size both in global/lumped terms and spatially distributed terms. In lumped terms, you could compare the size distribution for triggered landslides with the prior distribution. In the current case the prior is the observed size distribution for the study area but you could equally impose a uniform prior and assess the extent to the posterior matches the observed approaches both would be informative. In spatially distributed terms, the pattern of landslide size and its relationship to local conditions would be interesting and you would also be able to assess model performance with respect to landslides size by comparing the areas of predicted landslides that overlap observed landslides and (do they correlate? What is the form of the relationship?). Perhaps this type of analysis is reserved for a later study but it would fit nicely in the current paper.

It is an interesting point of back-calculating and validation of landslide size distribution. This is however, not in the scope of this paper. Regarding size of the paper and as suggested by the CC we would like to reserve this for a future publication and mention it explicitly in the outlook part of the paper (conclusion). A size-specific distribution is actually used in SlideforNet (https://www.ecorisq.org/slidefor-net-en), a lumped version of SlideforMap for practitioners.

Beyond these three major points I have several other questions that are more specific but less important. I don’t expect any of them to alter the primary messages of either the paper or the points I raise above but I hope they might be useful for the authors during revision. I do not understand the rationale behind some of the assumptions in SfM’s
Neglecting lateral earth pressure. It is true that active and passive earth pressure are maximised at some strain but neglecting them on this basis leaves two problems: a) you still need a treatment for the forces acting at the head and toe of the landslide; b) you need to apply the same criteria to root reinforcement since this is also maximised at some strain.

Thanks for pointing out that we did not do a good job in explaining the rationale behind our assumption. This will be specified in the revised version. In field experiments (Askarinejad et al., 2012) and numerical simulations (Cohen & Schwarz, 2017) it was shown that soil cohesion, active/passive earth pressure and root reinforcement are activated at different phases during soil replacement. a) We assume the phase where lateral root reinforcement in the tension crack is maximized to be the most stable, therefore we do our force balance here (as a sort of worst-case). Here we neglect passive earth pressure at the toe of the landslide because in that phase it is not usually activated (Cohen & Schwarz, 2017). b) This is the reason we only consider root reinforcement under tension (and not under compression, at this stage)

Neglecting soil cohesion on the sides. It seems inconsistent to apply root reinforcement but not soil cohesion on the lateral boundaries if you apply both on the base

We make the assumption that soil cohesion and root reinforcement are not additive along tension cracks at the lateral boundaries. With increasing displacement, roots exert an increasing root reinforcement until they break. Soil cohesion, however, only acts under very small displacement. At this point either one or the other acts. Only on the shear plane they are additive considering the effect of cohesion as component of the residual shear strength of the soil (Cohen et al., 2011).

We focus on the tension boundary of the landslide because it has been shown that roots do not contribute considerably to the maximum passive earth pressure forces. The passive earth pressure and compression reinforcement are activated at a later state in shallow landslide initiation than the tension forces (Schwarz et al., 2005).

All above will be specified better in the revised version of the paper.
Lateral root reinforcement acts only over the upslope half of the landslide’s perimeter (Eqn 3). I don’t see a justification for this and Schwarz et al., 2010 point out that it underestimates lateral reinforcement.

Landslide triggering happens in different phases. In our model, a first step is to decide on which triggering phase to focus on, i.e. which one to parameterize. As stated in the previous answer, we focus on the phase of maximum root mobilization. This will be made clearer in the revised version. In Schwarz et al., 2010, only trees along the landslide scarp (with the landslide already having occurred) are included. This results in an underestimation, which is addressed in their discussion. In SfM, we have a complete spatial field, so all trees are included. Upslope part of the landslide perimeter, not of the trees.

Lateral root reinforcement in Eqn. 9 is depth independent. This seems inconsistent with observed depth dependent rooting (density and size); and the depth dependence of basal reinforcement in SfM (Eqn 10).

The reviewer is right. An integration of the lateral root reinforcement over the soil depth will be included in a recomputation and revised version of the paper. Although we do not expect results to change much because most of the lateral root reinforcement is concentrated in the first 0.5 m of the soil depth (Vergani et al., 2016 and Gehring et al., 2019). The vast majority of our probabilistic landslides are deeper than 0.5 m.

Calculating root reinforcement using spatially averaged distance to trees within the Gamma function. Previous applications of the Gamma function (Eqn 9) appear to use it to predict root reinforcement at a known distance from the nearest tree (Moos et al., 2016). Given its nonlinearity, is it reasonable to use an average distance in Eqn 9 rather than evaluating Eqn 9 for the distribution of distances then averaging?

This is a good point. This was a parsimonious choice to reduce the computational effort; for a future paper, we will compute the difference between the lateral root reinforcement related to the average distance or the average root reinforcement of the full distribution and decide based on the results, which formulation to keep.

Variability

The form amplitude and spatial pattern of variability in material properties are all likely
important in defining landslide location and size (e.g. Bellugi et al., 2021). Representing this variability seems important. I would have liked to see more detail on your rationale for your choice of distribution form and spatial (de)correlation. I recognise that observations to inform this are sparse and these properties are not well known. The normal distribution has some specific problems that you grapple with but that others chose to avoid by using a log-normal (e.g. Griffiths et al., 2007). You deal with unphysical negative values by truncating, and claim these are rare but this places strict constraints on the variability that you can impose (small coefficients of variation for soil depth and cohesion in Table 5). In the absence of evidence to the contrary, a distribution that is limited to positive values (e.g. lognormal) would seem a more appropriate choice.

Thank you for pointing this out. We agree that the normal distribution has its limitations and that a lognormal distribution is a more appropriate choice. We will apply this in a recomputation of the results.

Soil depth variability is treated slightly differently (spatially de-correlated but slope dependent). I was unsure whether soil depths distribution was parameterised from observed landslide scar depths (L178) or using mean and standard deviation as parameters to optimise (Figure 7). The former seems problematic: landslides likely occur in deeper soils biasing the sample. Perhaps Eqn 7 was designed to account for this? However, I don’t understand why the coefficients on mu (1.35) and sigma1 (0.75) in Eqn 7 have these particular values. The second approach, tuning mean depth rather than setting it from observations seems more appealing to me and would also enable a comparison between model results and observed landslide depths, which would be a nice addition.

Mean and standard deviation of the soil depth are tuned parameters. In our results in Fig. 7, we used the calibrated values from Table 6 for the mean and standard deviation of soil depth. The parameters 1.35 and 0.75 are part of a correction factor for the sampled soil depth to the slope. This was approximated by a survival function related to the slope which was calibrated manually based on the inventory. We will adjust L178. We will point out that calibration to the slope angle and the values for mu (1.35) and sigma (0.75) are optional.

Hydrology

Your approach is exactly the same as that of SHALSTAB and SINMAP but is considerably different from Topmodel (Beven and Kirkby, 1979). All three use a topographic index to define hydrologically similar units. Topmodel uses these (with simple treatments for evaporation and infiltration) to simulate a time-varying catchment averaged response to a rainfall timeseries that can be mapped back onto the HSUs; the others simply solve for a single steady recharge rate (neglecting these processes). Even the topographic index (i.e. A/sin(B)) differs from that of Topmodel (which uses ln(A/tan(B))). This reflects differences in reference frame (the sin vs tan) and assumed conductivity profile (uniform vs exponential). I don’t disagree with the approach but I think it follows Montgomery and Dietrich (1994) and Pack et al. (1998) so it would be simpler to say that. If you wanted to give credit to earlier work then the TOPOG model of O’loughlin (1986) was behind the original derivation of SHALSTAB and the first introduction of a topographic index was by
Thank you for pointing this out with the detailed references. We will let go of the formulation of using TOPmodel or TOPmodel assumptions and give explicit credit to O’loughlin (1986) and Kirkby (1975).

Previous papers that apply this hydrological model do not claim that it is particularly well suited to slopes with macropore flow. Montgomery et al. (2002) highlight the importance of macropores and fractures (and a steep soil water characteristic curve) for hillslope hydrologic response but also recognise that “that rapid pore pressure response that controls slope instability […] is driven by vertical flow, not lateral flow” (Montgomery et al., 2004). There is general agreement that lateral flow (modelled here) strongly influences the pore pressure field antecedent to a burst of rain that could initiate a landslide (Iverson, 2000; Montgomery et al., 2002; 2004). This has important implications for the approach though because it implies that Q/T is an index for the ‘propensity for landsliding’ rather than a parameter to be calibrated within a complete hydrological treatment. This explains the apparent problem of predicted pore pressures independent of rainfall duration but observations that landslide triggering depends on both intensity and duration. Broad spatial patterns of pore pressure and instability should be well captured but triggering rainfall properties may not be. In fact, discussion of the influence of macropores on pore pressure tends to focus on the unpredictable localised pressure peaks associated with constrictions or terminations to macropores (e.g. Pierson, 1983; Montgomery et al. 2002). Even given these limitations I don’t think this is a bad model relative to the alternatives because it captures broad phreatic surface patterns and I’m convinced that the finer detail of these patterns is set by (unknown and perhaps unknowable) heterogeneity in material properties (e.g. macropores). If so, a more refined and expensive hydrological model may improve predictions of spatial pore pressure patterns very little.

Thanks for this concise summary. We will update our introduction and the justification of the model to better relate it to underlying assumptions and previous work.

Vertical flow driving is dependent on the situation in our opinion (as experiments cited in the paper by Askarinejad et al., (2012)). Moreover Montgomery and Dietrich (2004) argue that "lateral pore pressure diffusion does not appear to govern the site response time". They explain this inconsistency arguing that "the lateral flow response is dominated by the advective response", giving credits to the importance of lateral preferential flow in macropores. We agree with the CC that the process of building pore water pressure in soil during the triggering of shallow landslide is far from fully understood and we agree that the proposed approach is a good compromise to consider rainfall characteristics in slope stability modelling.

**Sensitivity Analysis**

As you point out parameter interaction makes it very difficult to infer parameter sensitivity from Figure 7 I think that may make it difficult to support some of your
assertions in L388-395 because you cannot guarantee that interactions are not masking other stronger sensitivities. For me the clearest example is the interaction between P and T (Table 6). Both are listed as uncertain parameters within the sensitivity analysis but only feature in pore pressure definition and only in that equation as the P/T ratio. As a result their inclusion as two separate variables in this analysis is likely to lead to severe equifinality (with high or low values will result in the same outcome as long as P/T is constant). Why not include the ratio of the two in your sensitivity analysis?

Thanks for this important point. Our analysis on the sensitivity of the parameters values could indeed be refined as suggested above. We will complete further analyses and decide how to revise the manuscript. We will add the P/T ratio to Table 6, Figure 7 and Figure 8. Subsequently we will do an analysis on the equifinality as suggested by CC.

Queries on equations:

1) I think there is a dimensional problem in either the first term of Eqn 3 or the second term of Eqn 4. Eqn 10 expresses \( R_{bas} \) as a function of \( R_{lat} \) so I think both should be either a force per unit length or a stress. If \( R_{lat} \) (in Eqn 9) is a stress then Eqn3 is dimensionally incorrect because the first term is a force per unit length and the second a force. The first term needs integrating over landslide depth. This could take the form \( \cos(s) H \) if you assume reinforcement is depth invariant. However, this would then be inconsistent with Eqn10, which assumes that root reinforcement declines with depth. On the other hand, if \( R_{lat} \) is a force per unit length (which might be more consistent with Moos et al (2016), Fig 3) then the problem may be more difficult to solve because the lateral depth integrated stress (N/m) is being applied across a basal area (m^2).

Thanks for pointing this out. Indeed as consistent with Moos et al., (2016), \( R_{lat} \) is in force per length (N/m). \( R_{bas} \) however, is in Pa in accordance with Gehring et al., (2019), equation 4. It uses a dimension correction factor. Under the assumption of root isotropy, the dimension correction factor will have a value of 1 m^(-1).

2) Are h and H measured in a vertical reference frame as indicated in Figure 2? If so then I think there is a \( \cos(s) \) missing from Eqn 12. The first \( \cos(s) \) converts vertical depth to slope normal thickness, the second converts phreatic surface thickness to pressure head (under assumptions of: uniform steady slope parallel seepage).

Thank you for pointing this out, we will include this in a revised computation and version.

3) Eqn 15 is incorrect because the original equation calculates DBH in cm from tree height in metres (Dorren, 2017) but you use DBH in metres (L292). I think Eqn 15 should be adjusted to \( 0.01H^{1.25} \).
You are right, thanks for pointing this out. We will adjust this in the revisions. In the computations this went correctly.

**LITERATURE**

Amishev, D., Basher, L. R., Phillips, C. J., Hill, S., Marden, M., Bloomberg, M., & Moore, J. R. (2014). *New forest management approaches to steep hills*. Ministry for Primary Industries.

Cohen, D., & Schwarz, M. (2017). Tree-root control of shallow landslides. *Earth Surface Dynamics, 5*(3), 451-477.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., ... Boehner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev., 8, 1991–2007.

Gehring, E., Conedera, M., Maringer, J., Giadrossich, F., Guastini, E., & Schwarz, M. (2019). Shallow landslide disposition in burnt European beech (Fagus sylvatica L.) forests. Scientific Reports, 9(1), 1–11. https://doi.org/10.1038/s41598-019-45073-7

Kirkby MJ. 1975. Hydrograph modelling strategies. In *Proc in Phys & Hum Geog*, Reel et al. (eds). Heinemann: London; 69–90.

Montgomery, D. R., & Dietrich, W. E. (1994). A physically based model for the topographic control on shallow landsliding. Water Resources Research, 30(4), 1153–1171. https://doi.org/10.1029/93WR02979

Moos, C., Bebi, P., Graf, F., Mattli, J., Rickli, C., & Schwarz, M. (2016). How does forest structure affect root reinforcement and susceptibility to shallow landslides? Earth Surface Processes and Landforms, 41(7), 951–960. https://doi.org/10.1002/esp.3887

O’loughlin, E.M., 1986. Prediction of surface saturation zones in natural catchments by topographic analysis. *WRR, 22*(5).

Schwarz, M., Preti, F., Giadrossich, F., Lehmann, P., & Or, D. (2010). Quantifying the role of vegetation in slope stability: A case study in Tuscany (Italy). Ecological Engineering,
Vergani, C., Schwarz, M., Soldati, M., Corda, A., Giadrossich, F., Chiaradia, E. A., ... Bassanelli, C. (2016). Root reinforcement dynamics in subalpine spruce forests following timber harvest: A case study in Canton Schwyz, Switzerland. Catena, 143, 275–288. https://doi.org/10.1016/j.catena.2016.03.038