Review of van Zadelhoff et al. for NHESS by David Milledge

Thank you to the authors for taking the time to respond to my comments, almost all of which have been addressed. However, I have one major outstanding concern. I think that the equations for basal and lateral reinforcement are still incorrect. It is essential that the authors address this prior to publication. I also have three, fairly minor concerns that can be very easily addressed:

1) The need for a clearer justification for the choice to neglect passive resistance;
2) Some statements in the paper have not been updated to reflect new results and are now incorrect;
3) Discussion of similarities between your 'no vegetation' parameterisation of your model and SINMAP.

My other concerns from the previous review have been addressed. Each of the remaining concerns has some history in previous reviews I have tried to summarise my concern and the history from previous reviews below.

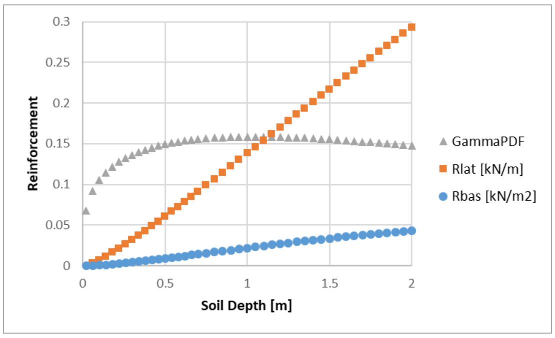
Major Comment: Representing basal root reinforcement as a function of lateral reinforcement is incorrect

It is now clear that you use a Gamma pdf and the additional equation is useful, you clarified in your response that you used numerical integration but should update the text to reflect this. However, my main original concern has not been addressed. Eqn10 calculates basal reinforcement as a function of lateral reinforcement. That is incorrect conceptually because if lateral reinforcement is a depth integration of reinforcement per unit area then basal reinforcement should simply be reinforcement per unit area at the basal depth. Your response to my previous comment suggests that you agree (you said *"The equation in essence is the same in both equation 8 and 9. Equation 8 however, applies the cumulative value (soil depth in range [0,x]), equation 9 applies the point value soil depth at [x]"*). However, your implementation is at odds with this. Eqn 4 in Gehring et al. (2019) is much more consistent with both your response and my expectation because they use of R_{max} rather than R_{lat}. Is it possible there is an error in Eqn 10 in the paper and it should be using R_{max}?

In addition to the conceptual problem, your approach introduces two errors:

1) The dimensional error that you have to correct with the unit coefficient k.
2) A relationship between basal reinforcement and depth that is at odds with observations of root density and reinforcement. Your group's previous observations (e.g. Schwarz et al., 2012) and those of other groups all suggest that reinforcement declines with depth, or at least does not increase. Your equations ensure that basal root reinforcement increases with depth. The graph below was generated using your gammaPDF parameters with an R_{max} ($\boldsymbol{R_{max}} = \boldsymbol{c}\,\boldsymbol{\Gamma}\left(\frac{\boldsymbol{D_{trees}}}{\boldsymbol{DBH\ D_{treesmax}}}\middle|\boldsymbol{\alpha_1}, \boldsymbol{\beta_1}\right)$) of 1 kN/m2 (to keep things simple).

I'm convinced that this representation is incorrect and that it is potentially important because in shallow soils, where basal reinforcement is particularly important, it will result in a very large underestimation of reinforcement.



My second concern having plotted the graph above is that the parameters of your gammaPDF implemented with Gehring et al.'s (2019) parameters implies that root density initially increases rapidly with depth while your own observations of the vertical distribution of roots in Switzerland and many observations from other groups all suggest that root biomass (and thus reinforcement) decays with depth, usually exponentially (see Dazio et al., 2018; Schwarz et al., 2012 and references therein: Abe and Iwamoto, 1990, Abe and Ziemer, 1991, Abernethy and Rutherfurd, 2001, Schmidt et al., 2001, Schenk and Jackson, 2002, Bischetti et al., 2005, Laio et al., 2006, Docker and Hubble, 2009). This suggests that your alpha2 and beta2 parameters are not physically reasonable.

Do you agree that gammaPDF values should be lower at 2 m depth than they are at 0.2 m depth?

A minor point, Table 1 would be improved by replacing 'literature' entries with the relevant references.

Minor Comment 1: The need for a clearer justification for your choice to neglect passive resistance.

I asked that you clarify the term: *"second stage of the activation phase"*; and provide citations to support your argument that passive resistance can be neglected in this phase.

In response you added the following text:

"The forces assumed in SlideforMap are typical for the second stage of the activation phase: the displacement at which lateral root reinforcement is maximised under tension along the tension crack and at which passive earth pressure, lateral root compression and lateral soil cohesion are assumed to not be fully mobilized (Cohen and Schwarz, 2017) and we choose to neglect these."

It is problematic to neglect the resistance on the basis that it is "not fully mobilised" because that gives no indication of its relative or absolute magnitude. It could mean (in relative terms) that the resistance in this phase is 90% of the maximum (fully mobilised) resistance; alternatively, even if the resistance were only 10% of the maximum this could still be larger than the resistance provided by roots. In fact, Cohen and Schwarz (2017) say: *"The largest force that contributes to slope stability is soil compression in the area above the landslide toe (p472)"*; and *"soil compression, due to its magnitude, dominates and controls the slope stability and its time to failure. (p472)"; their Fig 3 suggests passive earth pressure >5 kPa for most of your second phase displacement range (0.01-0.1 m)*. I don't think you can use this argument to justify your choice. You might instead neglect passive resistance because it is too difficult or complex or uncertain to model but if that is the case you should say so.

It would also be helpful to explain your choice to model the second activation phase rather than first or third. Why is the second phase the most important phase to model?

Minor Comment 2: Some statements in the paper have not been updated to reflect new results and are now incorrect.

For example you say *"The mean angle of internal friction shows a high variation, from a very low value for StA to close to the maximum tested value for Trub" (*L498-9*)*; but with the new results, variation is much reduced, Trub is far from maximum and StA is now the second largest value. On L527 you say *"This is unique for shallow landslide probability models"*; but this is contradicted in your discussion of PRIMULA as a probability model with a treatment of basal and lateral root reinforcement (L607-13).

Minor Comment 3: Recognising similarities between your 'no vegetation' model and SINMAP.

I suggested that *in the discussion of similarities to other models (Section 5.4) you explain that your 'no vegetation' case differs from SINMAP only in: 1) the form of the distributions that you use to sample c and phi (log-normal for you, uniform for SINMAP); and 2) the treatment of soil depth, which is spatially uniform in SINMAP and spatially variable for you. This means that the AUC values for 'no vegetation' are indicative of those that you would expect from SINMAP.* This was not addressed in your response but I think it is important that you do add this discussion because it has implications for the trade-off between model complexity and model skill. The no veg case represents a simple (low parameter) infinite slope model very similar to SINMAP that could be widely applied. If it works almost as well as the more complex model then that is an important finding worth reporting.

A summary of each of your vegetation scenario models, and what they mean for model structure would slot nicely into Section 3.4.3 'Vegetation parameter scenarios' (L425-30). Section 5.4 would be the place to add any interpretation of the performance of different parameter scenarios in terms of their implications for the value of different model structures i.e. what is gained by representing lateral reinforcement in general then what further gains are available when you use single tree detection.

In Section 5.6 (L601) you say: *"In comparison to more simple models based on infinite slope calculations (Pack et al., 1998; Montgomery and Dietrich, 1994, SINMAP,SHALTAB), SlideforMAP considers the effect of lateral root reinforcement".* This is the perfect place to point out that SfM in the 'no veg' case becomes an infinite slope calculation that does not consider the effect of lateral root reinforcement. You could then report performance of this version (which neglects lateral reinforcement) and comment on the performance improvement associated with representing lateral reinforcement.

My suggestions in relation to Minor Comment 3 are only suggestions though, not requirements, I'm happy for you to simply say no thankyou we don't want to do that.

Line specific comments:
L473: Why did "10 repetitions" of the global uniform veg produce such different mean AUC values (Table 8) to the AUC recovered during calibration (Table 7)? Global mean AUCs for each catchment (Table 8) are 4-7% lower than calibrated values (Table 7). The calibrated value is 1.8-3.9 standard deviations from the mean depending on catchment. Perhaps this is because the calibrated AUC takes the maximum from a large number of random realisations. However, in that case the calibrated AUCs (and associated ROC curves, Fig 8) are an inflated values, which are not representative and should not be reported elsewhere in the paper (e.g. L15, L636). The AUCs in Table 8 are much more representative and should be reported instead. You should consider entirely removing the AUC row from Table 7.

L475: "compared to the reference scenario used for calibration, the AUC remains almost unchanged or slightly increases for the vegetation based on single tree detection.": I have three points here: 1) this text appears to contradict later claims that single tree detection improves model performance (I think this problem will go away when you address my second point); 2) you shouldn't compare to the AUC for your reference because this is the maximum of a much larger number of trials (see previous comment) so the comparison is not a fair one; 3) Can you be more specific? This is a key result! it is worth adding a little more detail on the relative improvement for each site and on average. From a quick look, I think that in 2 of 3 cases AUC is increased by 1-4% in the other it is reduced by 4% and on average there is a 1% increase in AUC.

L549: "relative gain of 2%": It isn't clear what this comparison is referring to, which two vegetation parameterisations are being compared? How significant is the change? You can now comment on this using Table 9.

L551: "single tree detection is the best performing scenario": This is a key result. Can you give more detail here? What about StA and what about on average? How much better is it than the next best scenario? Is it significantly better than all the other scenarios or only a subset? Again this is a suggestion, not critical for publication.

Additional reference (not cited in paper)
Dazio, E., Conedera, M. and Schwarz, M., 2018. Impact of different chestnut coppice managements on root reinforcement and shallow landslide susceptibility. *Forest Ecology and Management*, *417*, pp.63-76.