REVIEWER1

The authors have carefully considered my comments and the issues I raised during the first round of review. I think that this revised version of the manuscript is acceptable for publication in HESS.
I invite the authors to consider only the following considerations, that could lead to some minor revision of the Discussion and, maybe, to modify something also in the Conclusions.

The comparison between the landslide predictions carried out with the (coarse-gridded) physically based model, with those obtained with a conceptual hydrological model (with finer resolution), shows that the latter still outperforms the first one, although in this new version the Authors exploit in a better way the outcome of the physically based model and obtain some different indications, compared to the former analyses.
In the Discussion section, they ascribe the limits of the physically based analysis mainly to the coarse resolution (of both the model itself and of input data about geomorphological and geotechnical parameters).
I agree with this intepretation, but I would like the authors to consider also another point (and add some comment in this respect, if they see my point).
They state that the coarse-gridded physically based model cannot reproduce the (lateral) fluxes between cells, as they are too distant from each other (the model grid is 12.5x12.5 km), as the high number of events during which nothing happens to soil moisture seems to demonstrate. Differently, the (conceptual) hydrological model, providing soil moisture at 0.5x0.5 km resolution, gives more valuable information to predict landslides.
What I want to stress here is that in initially unsaturated (shallow) soil covers, the prevailing direction of water fluxes is close to the orthogonal to the ground surface (e.g. Lu et al., 2011), owing to the top (atmosphere) and bottom (whatever it is) boundary conditions, that make the orthogonal hydraulic gradient much much larger than the one parallel to the slope (because all the verticals share more or less the same hydraulic conditions, and so there is very little gradient along the slope (which is at least one order of magnitude longer than the thickness of the soil cover). Only when saturation is reached somewhere within the slope cover, then, in that saturated part, the orthogonal hydraulic gradient becomes of the same order of the one parallel to the slope (depending on slope and bedrock inclination), and so lateral fluxes become significant (leading to subsurface runoff generation). However, I guess that most of the conditionally unstable slopes would have already failed before saturation was reached. I don't expect this picture to change significantly if the model was run over a 0.5x0.5 km gird instead of the 12.5x12.5 grid.
Far from saturation the only way infiltrating water can be drained out of the slope cover is either evapotranspiration (a too slow process over the time scale of 1 to 6 days of rainfall events) or drainage through the soil-bedrock interface, which becomes the most delicate point of the physically based model, about which the authors do not give any information to the reader.
Although I did not look in the cited paper where the PREVAH model is described, I expect that water exchange mechanisms between the three storage modules are somehow introduced in such model, and if the model is somehow calibrated in order to provide reliable results, these mechanisms consider the water exchange between the two upper storage modules and the lower one. This could explain why the dynamics of soil saturation estimated by the PREVAH model better matches with the effects of single rainfall events.
So, at least this is my opinion, the physically based predictions are limited not only by resolution and/or accuracy of input data issues, but also by the unsuitability of thechosen model to correctly assess the effects of one of the (major) processes controlling the dynamics of soil moisture in the unsaturated zone, i.e. the leakage towards the underlying saturated zone.

N. Lu, B.S. Kaya, J.W. Godt (2011). Direction of unsaturated flow in a homogeneous and isotropic hillslope. Water Resources Research 47(2), https://doi.org/10.1029/2010WR010003

We thank the reviewer for the positive comments and the very interesting points raised.
We agree that vertical infiltration is dominant in unsaturated soils because the vertical pressure gradients and fluxes are greater than the lateral ones, typically during landslide events (Iverson 2000), but we still believe that lateral flow is essential for determining realistic initial conditions for slope failure initiation on hillslopes with some topography (as we showed in Leonarduzzi et el., 2021). In fact, we do not believe that the limiting factor of having the physics-based model run at such coarse resolution is (only) the large distance between cells, but much more the smoothing of topography that destroys possible lateral gradients and of the meteorological forcing. It is hard to conclude this from the experiments carried out here, but we believe that increasing the resolution will indeed significantly improve the representation of water flow, forcing, and consequently the estimates of soil moisture. Furthermore, PREVAH (as most hydrological models) is calibrated for runoff, so it is not explicitly designed for realistically reproducing subsurface storages.

The second point raised, that the physically based model predictions may also be inaccurate because of missing the leakage process in deep soil, is a very good one, and we have added it in the discussion as a possible explanation. Indeed, most models simply assume a homogeneous soil of a certain thickness, or a multilayered soil with prescribed hydraulic conductivity (like we did). However, reality is much more complicated at the interface of the soil and weathered bedrock or saturated layer, where vertical flows can be affected by fractures, preferential flow paths, etc. The conceptual model also does not reproduce exactly these processes, but because it has several soil storage layers, it perhaps manages to get a better soil moisture distribution for this reason.

REVIEWER 2

This paper studies the prediction of landslide initiation by adding hydrological information to the well-known meteorological thresholds most frequently used and compares probabilistic and physically based modelling approaches for inclusion of antecedent soil wetness state (or water table) in the prediction. This review is on the revised paper. I have not been involved in the first round of reviews.

This paper is timely and contributes to an important research question in landslide research: can landslide forecasting on regional scale be improved by including hydrological information compared to traditional meteorological threshold based forecasting. And if so, how? The paper contains a huge amount of work and data. The methods are sound and clearly described. It studies the obvious question to what extend landslide forecasting can make use of existing regional hydrological model outputs. It tests this using two different regional models with different concepts and resolutions. The underlying work is in my opinion a creative and innovative contribution to this research field. The paper is well structured and written.

Overall, the study concludes that the contribution of the output of an existing large-scale hydrological model with coarse spatial resolution combined with infinite slope FoS calculation in the current state does not lead to an improved landslide forecasting. However, adding the hydrological information from a water balance model (with higher spatial resolution) modestly improve landslide forecasting in Switzerland. It shows that antecedent soil moisture information improves the landslide prediction using the probabilistic framework, but in a quite rudimentary way: by splitting the data sets in dry and wet antecedent conditions (just like proposed frequently before in literature, such as Figure 5.2 published in the Sidle-Ochiai, 2006).

Looking at the first round of reviews, I am of the opinion the authors did a profound job addressing the valuable comments and improved the paper significantly. The fact that this study uses existing models and combines those in landslide forecasting does not mean it is not novel, on the contrary.

I think the paper can be accepted for publication in Hess with some minor revisions.

We thank the reviewer for the positive feedback and the insightful comments.

Comments

- The authors sometimes use words or abbreviations of words in formula's and in figure caption and axis titles. I personally find those difficult to read and suggest to use 1 symbol for a variable or parameter and subscripts (e.g. x-axis of Fig 4, 5).

Because we are considering both temporal average in the cell and max-min during the specific event, we believe changing this to subscripts will possibly make it simpler but easily to misunderstand.

- L256: does a FoS needs to be 'exceptionally' low? I guess just a certain degree lower could be sufficient? And Figure 4 shows that a bit lower starting point is sufficient.

Yes, it doesn't need to be much lower during triggering events, what we meant to say is that we were trying to find a clear distinction between triggering and non-triggering events, with the FoS being smaller during the former. We removed "exceptionally".

- L278: Would it not be easier for the reader if you use one dominant terminology: False alarm, misses, TP, TN (and later on T, above/below, NT_above/below). So use TP, TN, FP, FN (and add false alarm or missed alarms in parenthesis to it a few times)? Same for Figure 6.

We agree with this comment and will be consistent in the revised manuscript.

- L298: layout exponent, not in superscript

We have corrected this.

- L300: Figure 8: Honestly, is the TSS really that much different for the different graphs? I would argue you can use all these graphs to split the data set. The entire range of the TSS as function of antecedent wetness is 0.1…. that is not a lot, is it?

This is a commonly raised question. The differences in TSS are generally very small, so a difference of 0.1 is, relatively speaking, not that small. We agree that using a threshold of antecedent moisture over any of the antecedent periods considered would not have made a large difference, but we carried out these experiments exactly to a) check whether this was the case and b) optimize the mean antecedent saturation threshold value.

- L323: vegetation cohesion is not a correct terminology. It is soil cohesion and apparent root cohesion: combined used in the infinite slope model.

We have corrected this.

- L324: for shallow slip surfaces (<2m) cohesion is a very sensitive parameter, by definition. Especially if apparent root cohesion is added.

We agree.

- L346: I challenge the authors with this statement. Why would a physically based model per definition (or theoretically) be superior? The so-called physical laws do not apply to the scale of the model but have another Representative Elemental Volume. Equally, could we not argue that on the regional scale we need another set of physical laws. So why can we not say that a physically conceptual model will be theoretically superior to describe the hydrological system on regional scale better than a bottom-up physically based model applied to the 'wrong' scale? At least you should address the scale issue related to the used 'physics'. You could elaborate a bit more on this.

That is exactly the point we are trying to make. We are not trying to draw conclusions that go beyond what we are showing here, but we do say that while the physics-based model is in principle closer to the process representation, at this resolution it is simply incapable of adequately reproducing the soil moisture dynamics that are required for the landslides modeling.

- L360: Maybe the authors can reflect a little why using regional hydrological models have only modest improvement to landslide forecasts and how that compares to hydrometeorological thresholds for landslides using in-situ hydrological (soil moisture) measurements (in Switzerland).

To the authors knowledge nobody has yet combined in situ soil moisture measurements with rainfall observations (hydrometeorological thresholds) in Switzerland, so it is hard to comment on this. The only contribution is that presented in Wicki et al. (2021), where modeled and measured soil moisture is compared for landslide predictions, but soil moisture thresholds are considered, without including rainfall.

- L367: I am a bit surprised to read the first conclusion. The first conclusion is about the effect of cohesion (in your case soil cohesion and apparent root cohesion) or not. I do not consider this the first and most important conclusion. As mentioned above as well, for shallow landslides cohesion can be the dominant resisting force, so putting that to zero is not surprisingly changing the areas where landslides can take place. I suggest to make this not the first conclusion but rather the 3rd or so.

The order of the conclusion matches the order of the research questions in the introduction and the of the results presentation, not the relative importance of them.

- L376: This conclusion is not fully justified, or at least not telling the complete story. Another way of framing this conclusion could be stressing that in your analysis 65% of the country is US (35% when C=0 kpa) That leaves 35% (or 65%) to be potentially 'triggerable'. This is how I read this result.

This of course is true, but the point of this analysis was not really to assess landslide susceptibility, but to see where hydrology can play a role if we choose to follow the infinite slope approach. The idea is that this could help a) understand if the infinite slope approach is suitable at all (by comparing to landslide observations), and b) help constraint the hydrological simulation to a much smaller portion of the country (as there is no need to get saturation estimates in UU or US regions).

REVIEWER 3

There are some minor typing corrections and some further explanations required, as indicated in the manuscript.

Technical comments:

1. Page 12, line 277: how did you obtain this equation and the value of TSS.

We briefly explain the procedure in lines 182-184 and reference to our previous work where it is explained more in details. We describe the power law equations as $E=a*D^b$ and then test all possible combinations of a and b parameters and pick the one maximizing the TSS.

2. Page 13, line 280: this conclusion is not always true for six days antecedent saturation before landslides in all plots; i.e., less antecedent saturation governed the miss (T below). Please explain this discrepancy.

The reason for this discrepancy is that there are so few triggering events below the threshold of duration 6d that the mean antecedent saturation is statistically not meaningful. In fact, the red dashed line in the lower right panel of Figure 6 drops to basically zero for duration of 6d. We have added a comment to this in the revised manuscript.

3. Page 14, Figure 6: the legend used in all plots should be changed to True positive, miss, false alarm, and true negative.

We have changed this in the revised manuscript.