We thank the reviewer for the review and the constructive comments. We address here all the points raised, and we indicate how we will take care of them in the revision.

The paper investigates the use of (very) large scale hydrological modeling to improve the prediction of shallow landslide occurrence throughout Switzerland, compared to the typical approach based on the statistical analysis of triggering precipitation alone. This is a quite actual topic, which can be of interest for the readership of HESS, as most landslide early warning systems still rely only on precipitation information, while there is physical and operational evidence that including hydrologic information may be useful in many cases.

The paper is well organized and the English language pretty good, although the choice of the scatter plots adopted for presenting the results may result a bit awkward (although quite synoptic, which is a good point).

We agree that the scatter plots (e.g. Fig 4 and 6) are not easily readable. The idea was to look at how “exceptional” triggering days are in terms of antecedent (saturation or FoS) and triggering (rainfall) conditions. We will improve the description and consider simplifying them.

However, I find some issues in the adopted modeling approaches, which somehow affect also the results and the conclusions, so I believe that major revisions are needed before reevaluating the manuscript for possible publication.

Specifically, all the paper deals with the comparison of a “physically-based” hydrological model, run over a coarse spatial grid and coupled with a simplified slope equilibrium equation based on the infinite slope hypothesis so to end with an assessment of a safety factor value at any point of the grid, with a purely probabilistic evaluation of the coupled effects of slope conditions in landslide and non-landslide days, carried out by estimating soil degree of saturation with a conceptual hydrological model run at a much finer spatial scale. The obtained results indicate that the “physically-based” approach is largely outperformed by the probabilistic, and the discussion ascribes this outcome mostly to issues related to the coarse resolution (i.e. wrong local estimates of soil depth, slope inclination, soil mechanical properties, slope hydraulic response to precipitations, and so on). All these discussion points are clearly valid and acceptable, but I believe that the Authors should more deeply describe, discuss and comment the limitations of the model that they consider as “physically-based”.

We refer to these approach as physically based according to the classical definition in this field, insofar hydrological and geotechnical modelling is concerned. We focused on the issues associated with the resolution, because they are definitely key here. In fact, from a process description point of view for surface and subsurface water fluxes, the physically-based framework of TerrSysMP should be clearly superior to the conceptual PREVAH, and the resolution must be a key difference. Solving water flow equations over a grid of 12.5 km is simply too coarse to get the fluxes right, regardless of the accuracy of the physics behind the flow equations. Nevertheless, there are of course other limitations also associated with all the components of the modeling framework, we will add a comment on these.

I understand that the Authors probably mean that with such a modeling approach they assess landslide occurrence with an equilibrium equation, that is the application of a physical principle. However, although no detail is provided about the characteristics of the infiltration model which provides the water table depth for the application of the equilibrium equation, I have the feeling that it may be not completely physically based. In fact, while dealing with shallow landslides, which occur in initially unsaturated soil covers (as the Authors indeed notice in the Discussion section), it is *assumed that the infiltration process results in the building of a water table at some depth in the soil, which is not necessarily the case (it strongly depends on the assumed boundary condition at the base of the soil cover), and which seems more a conceptualization of the effects of infiltration, rather than the result of a physically-based model of rainfall infiltration process. The adopted expression (1) of the factor of safety (and the obtained results, as well as the discussion about them) seem to be deeply affected by this conceptualization.*

The soil hydrological model in TerrSysMP solves 3D Richards equation in the subsurface and computes overland flow in a fully coupled manner with a kinematic wave approximation. The development of a saturated layer is possible at any depth in the soil depending on the soil layering. We will add a description of the hydrological model in section 2.1.3. The second point raised by the reviewer, that wtd might not be the most adequate for the FoS calculation is a fair point. We had tested saturation from TerrSysMP unsuccessfully, but in the process of further exploring to answer this point raised, we solved an issue that was impacting the results. While the results are still not superior to those of the saturation estimated by the conceptual hydrological model PREVAH, they are more informative than the wtd results (e.g. see Figure 1 here below). We will explore the potential of TerrSysMP further and add the results in the manuscript together with a 1:1 comparison with PREVAH's saturation.

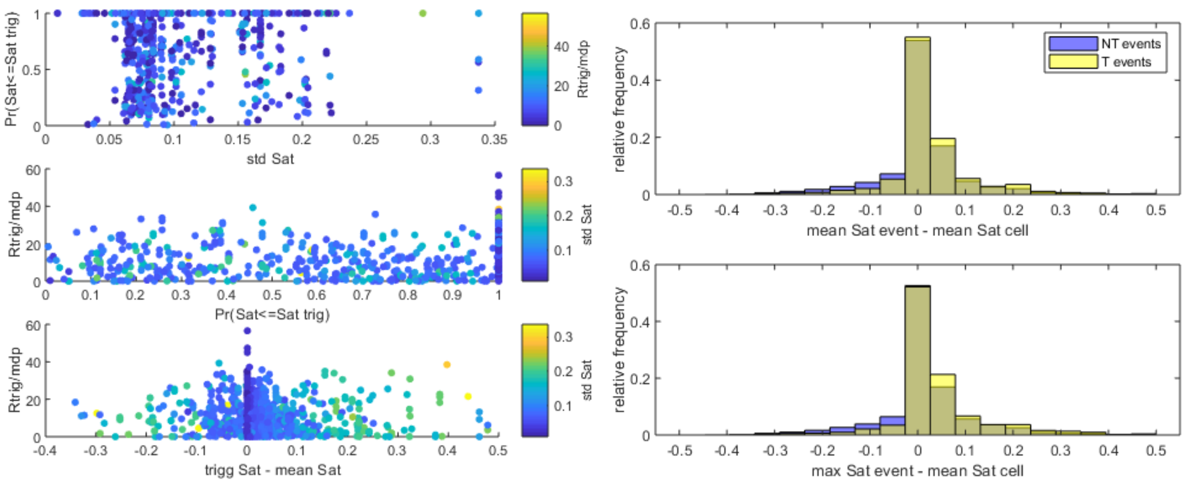

Figure 1. Right, the scatter-plot of different combinations of the rainfall and saturation (from TerrSysMP, as the average saturation in the first two soil layers, 60cm) properties for all landslides (each point corresponds to a landslide): the probability of the saturation in the cell being smaller than the value on the day of the landslide (Pr(Sat<=Sat trig)), the standard deviation in time of the saturation in the cell (std Sat), the ratio between the triggering rainfall intensity (rainfall intensity on the day of the landslide) and the cell mean daily precipitation (Rtrig/mdp), and the difference between the triggering saturation and the temporal mean saturation of the cell (trigg Sat - mean Sat). Left, the histograms of the mean (top) and maximum (bottom) saturation estimated by TerrSysMP during triggering and non-triggering rainfall events, combining spatial (i.e. differences between landslide locations) and temporal (i.e. differences between events in the cells) differences.

1. When the soil is considered dry and cohesionless (h=0 in equation (1)), FoS reduces to tan(fi)/tan(beta), which implies that soil depth is ineffective and that everything depends on the quality of your topographic map (beta varies much more than fi, which not surprisingly remains always not far from 30 for the kind of soils you may have in mountain environment).

2. When root-cohesion is introduced (by the way, another conceptualization), considering dry slopes with inclination larger than 30, you can easily see that it mostly results FoS>tan(fi)/tan(beta)+0.1c, so that even with the smallest hypothesized cohesion (5 kPa), FoS can be smaller than one only for slopes more inclined than 50.

*3. When saturated soil cover (h=d) is considered without root cohesion, it is FoS=(ggw)/g*tan(fi)/tan(beta), that, with the values of g that you assume for the soil (seemingly between 1.2 and 1.6, with gwïA¸A˘ 1) leads to stability possible only for inclinations smaller than 12, independent of soil depth.*

*4. If we introduce root cohesion when h=d, you get again that for slopes with inclination above 30 some 0.1c is summed up to the previous expression of FoS, that is FoSïA¸A˘ 0.1c+(g-gw)/g*tan(fi)/tan(beta), so that only when cohesion is the smallest (c=5 kPa) you may get some slope inclinations for which stability depends on the value of the water table h.*

The referee is describing exactly what we called the unconditionally stable and unstable conditions with the given datasets we used. Perhaps we were not clear in stating that and we will rephrase this part of the manuscript. Our aim was to actually quantify that the range, as the area where hydrology matters and instability is possible, and show that it is restricted to a rather small part of the country when following the FoS approach.

This given, my overall impression is that all the results from the "physically-based" modeling, namely all the considerations about (un)conditional (un)stable situations, and their comparison with the landslide inventory are strongly affected by the weakness of the model, before than by the issues related to the coarse modeling grid.

We agree completely with this statement. In fact, the purpose of looking at (un)conditional (un)stable regions (i.e. susceptibility) it's exactly to split the impact of the hydrology (considered in the dynamic comparison instead) and look at the geotechnical model (infinite slope approach) only. The results concerning part (i.e. Figure 3 and Table 1 in the manuscript) are completely independent from the hydrological component and therefore its limitations are only due to the model assumptions and the uncertain input parameters. At the same time, the dynamic soil wetness is affected by the coarse grid resolution, so ultimately it is the combination of the inadequacy of the infinite slope geotechnical model together with the soil wetness state that limits predictions. We will make the separation between hydrological and geotechnical component clearer in the revised manuscript.

I mention a few points that I believe are worth some discussion: (i) what is the meaning of gamma in equation (1)? This value should change according to soil saturation, and the assumed values between 12 and 16 kN/m3 seem rather to refer to some average field condition (this certainly has an effect on the predicted values of FoS);

The correct procedure would be to compute it based on the soil bulk density (ρ_s), water unit weight (γ_w), and the void ratio. The spatially variables parameters are ρ_s and the void ratio. While a map is available for the soil bulk density, the void ratio could only be estimated given the soil type. As for the friction angle, the resulting values would be very homogeneous over the country. We therefore simplified by taking the dry soil unit weight value instead. This is a strong simplification, but highlights one of the key messages of this paper, which is that there are strong limitations associated with a physically based approach that should be taken into account when designing a landslide prediction system. We will stress this message and add the explanation of the equation's parameter.

(ii) to what extent the assumption of the building of a water table is acceptable and consistent with the geomorphological characteristics of the studied alpine slopes (i.e. type of soil and type of bedrock)? (iii) is groundwater table (likely much deeper than the shallow soil covers of interest for the study, as the Authors themselves observe at lines 270-271) an appropriate variable to be chosen for the purpose of this study about shallow landslides? (iv) I guess that TerrSysMP model offers also soil

moisture data, so why did you choose groundwater table for your analyses? Concluding on this point, I still believe that the attempt to exploit the information available from a model like TerrSysMP for the sake of predicting landslides is a valuable task, and that it merits to be investigated. But it seems to me that this could be made with more care than it is in this study.

The answer to this comment can be found above. We will add information about the results considering saturation provided by TerrSysMP also in the revised manuscript (similarly to the plots here above) and a comparison to PREVAH's saturation results.

On the other hand, there is the conceptual hydrologic model and the use of estimated soil moisture with a probabilistic approach to improve landslide assessment carried out with empirical precipitation thresholds. While this part is more straightforward, there is still a major point that should be clarified. Your aim is to investigate the potential of soil moisture prior the onset of triggering rainfall to improve empirical thresholds. Despite this, from figures 6 and 7 it seems that you never consider this information, as only saturation on the day of the landslide, maximum or mean saturation during an event, and general statistics of the saturation in the cells are calculated. The discussion of the moisture conditions prior the event is limited to graphs of fig. 8, considering mean saturation for 5-60 days long periods preceding rainfall events. Some discussion of the graphs would be worth. For instance: the 5 and 10 days averages seem to be the best choice to correct false alarms (red line well below the others); long events (6 days) seem to lose memory of the effects of initial conditions on missed alarms (all yellow lines drop down for 6 days, while they are above all other lines for shorter event durations).

The referee is correct in this statement. The point of Figure 6 was indeed to 1) evaluate if the saturation was exceptionally high on landslide days, and 2) see if that related to the triggering rainfall intensity. We considered accounting for saturation over longer periods of time in Figure 6 of the manuscript, but decided not to for two reasons: first, to facilitate comparison to the equivalent Figure 4 of the physically based approach, second, because it would complicate the plot even more. We will consider adding in the revised manuscript or its supplementary material a version of Figure 7, where the saturation over N days prior to the rainfall event is considered. We will also add some more discussion of the different windows prior to landsliding and the effects they may have in rainfall ID prediction in the revised manuscript.

Instead, in the paper only the brief sentence at lines 303-304 is dedicated to the possibility of building hydrometeorological thresholds, which are just said to be uncapable of improving the performance of precipitation thresholds without any information. I think that much more discussion and data should be presented to the reader, as the effects of prior soil moisture is all in all the focus of the paper.

What we meant is that replacing duration of the intensity-duration threshold with e.g. saturation (leading to hydrometeorological thresholds) actually lead to worsen performances, while introducting a saturation threshold which splits the event in high and low antecedent saturation followed by two individual ED (cumulative rainfall vs duration) thresholds shows an improvement compared to a unique rainfall threshold. We will add more information about this attempt and the performances obtained.

In addition to these two major issues, you can find some remarks and comments as annotations in the attached file.

We report here few comments on the annotations in the pdf:

- We consider indeed predisposing factors as static, as they generally change over very long timeframe. This is very different from the cause (hydrological conditions) and trigger factors which instead we consider to be dynamic. A description of the latter is provided in Bogaard and Greco 2018. We will specify this.
- Fig. 1: a different colormap was suggested to allow better visualization of the friction angle values. Unfortunately, there are really just 2 unique values for the entire country which are visible in yellow and orange.
- We agree that ignoring the increase of soil strength due to soil suction might explain some the results and most importantly the existence of the unconditionally unstable areas. We will add this consideration.

Fig. 2: Looking at the color maps, it seems that while the slope and elevation dependent depth models result in thinner soil cover in the mountains, the lin. diffusion works in the opposite way around. Also the zoomed pixel apparently shows that the blue-colored (nearly zero depth) lines of the first two models become yellowish in the third model (nearly 2 meters depth). How can you explain this discrepancy?

Actually the top of the mountains has thin soil cover also in the case of the linear diffusion model (the mountain crests appear like “blue rivers”). Right next to the ridge, the soil often becomes thicker due to the dependence to the second derivative of elevation (curvature).

Bogaard, T. and Greco, R.: Invited perspectives: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: proposing hydro-meteorological thresholds, Nat. Hazards Earth Syst. Sci., 18, 31–39, https://doi.org/10.5194/nhess-18-31-2018, 2018.