Comment on hess-2021-310
Bettina Schaefli (Referee)

I read the manuscript and the review by reviewer 1. I completely agree with his point 1 that the literature review is incomplete. For each investigated modification, we would need a reference for who has done this before. Besides a review of how others tried to improve the degree-day method, we also need reference to pattern calibration in hydrology, i.e. for calibration on other patterns (e.g. evaporation).

I furthermore believe that something went wrong with the subsection numbers. It would be good to use classical sections (intro, case study, method, results, discussion, conclusion) and to avoid mini-subsections.

Overall, the idea to calibrate the simple snow routine (composed of an accumulation routine and a degree-day melt routine) on single-day snow image is indeed interesting. The results show that a calibration on a single snow image gives as good results as calibrating the snow routine directly on streamflow (the reference HBV). And furthermore, calibrated parameter values for different years are relatively stable.

While it can in principle be interesting to test different snow accumulation and melt approaches, I am not convinced that the presented approach is able to provide any valuable insights. The reason is that the provided model performance numbers do not tell us whether one model is really better than another one; what does the increase of NSE by 0.002 actually mean? Even an increase from 0.82 to 0.85, is this significant?

The advantage of calibrating the snow parameters directly on snow images as opposed to calibrating on discharge (which is still required to calibrate the other model parameters) is not clearly discussed.
There are only three case studies, on reflecting intermittent snow, the other one reflecting the typical build up of a seasonal snow cover. Is calibration on snow of a particular interest in one case or the other?

Finally, I believe something went also wrong with the choice of the Swiss catchment. I do not know which outlet you have chosen, but it seems to be (given the high discharge) after the outflow of a large lake in Luzern, which means that you emulate the behavior of the lake through the hydrological model. Any effect of the snow cover / melt will be strongly smoothed out by the lake volume.

For all above reasons, I cannot recommend a revision of this paper for publication in HESS. The material could be turned into an interesting contribution but would require substantial work, which I believe goes beyond major revisions.

**Detailed comments:**

- the water balance equation of the snow cover is missing; how does liquid precip enter the equation? ie. is liquid water immediately added to the snowpack outflow?

- melt does not increase because of liquid water falling on the snowpack, this is a physical misconception; in wet conditions (rainfall or high relative humidity), the heat transfer from the atmosphere to the snowpack is higher than in dry conditions; it is this heat transfer that increases the melt and not the advected heat via rainfall entering the snowpack (which is very small); and of course the liquid water content of the snowpack increases with incoming rain but we do not know how the liquid water content of the snow cover is modelled here (point above)

- list of parameters / variables are unusual, I would put in text format

- we need information on the used catchments (classical catchment characteristics, including location, size, etc).

- remove unnecessary digits in the tables

- how to you define winter? what are snow days? with snowfall or snow on the ground?
Any conceptual water-streamflow transformation model has to be calibrated with the water input. If you change the input, you have to recalibrate the other model parameters. Here, the input with the different snow models is probably only marginally different, but this should be mentioned anyway; and if the input is only marginally different, how can we conclude that the snow models lead to different performance.

- avoid multi-letter variable or parameter names

- NSE values for precipitation against elevation are not really interesting; NSE depends on the underlying signal seasonality, which varies with elevation and with region

- reference for Residual Kriging?

- time frame of simulations, which time period is used for the precip and temperature stations

- did you account for gauge undercatch for precip measurements during winter (snow)? If not, this will most likely strongly underestimate snowfall

- snow season in Switzerland can start in September but only at very high elevations, where melt season continues in July

- I would not discuss model updating and (real-time) forecasting; this is a very different topic and would require more references and in-depth discussion;