Response to Reviewer #2

We would like to extend our thanks to Bettina Schaefli for taking her time to thoroughly read the paper and give her critical remarks.

While there are many valid points in her review, there are quite a few points where we do not agree.

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).

This is statement related to fundamental problems of present publications. The second author of the paper serves himself as an editor and has the opinion that the introduction of the papers increased significantly in the past 20 years. The introductions are becoming medium quality mini review papers, which are unimportant for people who are working on the topic, and of little use for those who are unfamiliar with the topic. They serve mainly for the increase of citations.

The purpose of scientific publications is to share new ideas and experiences with fellow scientists. Review papers are an exception, but they should be written by experts with sufficient experience, and not by PhD students. The review process should focus on the novelty and applicability presented in the paper. Due to the present evaluation of scientists, it is unfortunately not likely that this tendency to fill a paper with a large number of references mainly in the introduction will change, as promotions depend on the number of papers and citations. We will however, follow the suggestions of the reviewers to rewrite an introduction with extensive references, but we wanted to express our discomfort.

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.

This can be done, we will reformulate the sections and avoid the sub-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.

These are our main findings.

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?*

In this case the calibration of the snow model was not done with the help of the hydrological model, but instead independently on satellite images. If one would calibrate the snow model using the hydrological model the uncertainty would increase due to the fact, that other parameters of the model may compensate for problems in the snow module. It is always advantageous to try to calibrate submodels independently and on other quantities. This way the model parameters are likely to become good because of the right reasons. This will also contribute to the reduction of the equifinality of the model parameters.

Further snowmelt models are not only important for hydrological modelling. For example the estimation of snow amounts are very important for water management. The suggested model could be inverted to provide estimates of snow amounts in remote regions.

An increase of the NS value of the hydrological model was not the goal, it is thought to show that the independently achieved snow module is good enough to become an non-discharge calibrated part of the hydrological model. The difference in the NS could be significant. It is not the difference which decides on the significance of the change of an NS value. As an analogous example consider the correlation coefficient: In a normal case with a sample size of 365, the increase of a correlation from 0.85 to 0.89 or from 0.94 to 0.96 is highly significant while a seemingly much higher increase of the correlation of 0.25 to 0.35 is not significant (all on 95 % significance level).

$$z = \frac{\ln \frac{1+\rho_1}{1-\rho_1} - \ln \frac{1+\rho_2}{1-\rho_2}}{2\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

For the NS values the idea is similar, but unfortunately the calculation of significance of the change of the NS value would require independence of the errors which is not the case. Further, there is no good model which could describe the dependence of the model errors. Treating this topic goes far beyond the scope of this paper. But for sure, one should not decide on the significance whether two NS values are different or should not be done only on the basis of difference of the NS values.

Fisher, R. A. (1921). On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample. Metron, 1, 3-32.

The advantage of calibrating the snow parameters directly on snow images as opposed to calibratinjg on discharge (which is still required to calibrate the other model parameters) is not clearly discussed.

A discussion on this will be added to the updated version of the paper. As mentioned above, the advantage of independent calibration of parts of the model contributes to the reduction of uncertainty and to the identification of modeling problems.

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?

Calibrating on snow cover is interesting in both cases. For intermittent snow cover it has direct influence on discharge, while for build up and melting it can contribute to the estimation of snow amounts.

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.

Thank you for the hint - we will use another catchment.

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.

This is a very disappointing statement as we believe that we presented a new and useful method, which could be useful for others, and your critics are mainly concerning the presentation, and very little additional computations are needed. Thus the revision of the paper could be done in a reasonable time.

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?

The snow-cover is only used for calibrating the snow models. Considering a grid cell, if the cell has snow water equivalent (SWE) greater than 0.5mm, it is reclassified as '1' and '0' if not. This would be our modeled binary snow cover distribution. The model is then calibrated using the MODIS snow cover distribution independently (also binary '1' for cells with snow and '0' for cells without) as the reference. The liquid water is not considered in the snow models. However, this is immediately added to the melt water outflow in the truncated HBV model.

- 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 atmospher 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 snowmpack

(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)

Thank you for the suggestion. We agree that the melt induced by the latent heat carried by the rain entering the snowpack, albeit less, contributes to melt. This section will be slightly rephrased. Please refer above for the last query.

- list of paramters / variables are unusal, I would put in text format

The feedback is well-noted.

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

This will be done in the revision.

- remove unnecessary digits in the tables

The digits will be rounded in the tables in the revised manuscript.

- how to you define winter? what are snow days? with snowfall or snow on the ground?

In this study, we have assumed winter as a mean snow-season for the regions. The winter season was considered from October to April in Baden-Wuerttemberg and September to June for Switzerland. 'Snow days' are defined as the days with available snow (above 0.5 mm SWE) in the regions till it disappears. This will be defined clearly in the revision.

-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.

As already mentioned above, the calibration of the snow model was not done with the help of the hydrological model, but instead independently on satellite images. However, we will re-calibrate the truncated HBV in the revision with the melt from snow-models as the standalone inputs. The results will then be discussed in the revised manuscript.

- avoid multi-letter variable or parameter names

The feedback is well noted. The letters were added to provide a clear definition of which variable is being used in the models to avoid confusion, as the models take different inputs.

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

The cross validation of temperature and precipitation is not a novel finding. It was only included to partly quantify the possible error of the input data. We intend to shorten this part of the paper.

- *reference for Residual Kriging?*

This will be added in the revised document.

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

For Baden-Wuerttemberg, the time period used was from 2010 to 2015 and for Switzerland it was from 2010 to 2018. This will be added in the revised document.

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

The gauge undercatch for precipitation measurements during winter was not accounted for. We could try including a correction factor in the revision.

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

Thank you for this suggestion. We have focused on the period from September to June for Switzerland to demarcate the 'snow' season.

- *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;*

Well noted. This will be revised.