General Response

We would like to thank both reviewers for taking their time to thoroughly review our paper and providing crucial feedback. We have streamlined our manuscript in the new revised version with detailed review of similar approaches carried out in the past. As per the reviewers' suggestions, we have tried to formulate a better narrative to the study. The review, we believe, has helped to improve the manuscript.

Three more catchments were added to analyse the efficacy of our approach. We have further discussed about how independent calibration helps in improving the snow-model performance thereby reducing the uncertainty related to snow melt simulation as compared to the hydrological model simulations, and how it can add value to discharge simulations.

Response to Reviewer #1

We thank Juraj Parajka for taking his time to carefully read the paper and providing critical remarks.

General comments

This study presents an approach for calibrating different degree-day snowmelt approaches by using MODIS snow cover data. The second aim is to examine different degree-day variants for snowmelt simulations and calibrate or validate them using satellite snow cover data. The approach is tested in two regions (Baden-Württemberg in Germany and Switzerland). The results indicate a slight increase in overall NSE runoff performance and a better NSE performance during the winter period.

I read with interest the manuscript because we did numerous similar experiments in the past (and recently). I have to say that the manuscript presents some interesting and novel experiments, but as a whole, it is not ready for publication in its current form. The main reasons for such assessment are:

The Introduction section needs to be improved. In its current form, it is not specifically presenting which approaches are already available, what the research gaps are and how this research goes beyond existing studies? There are numerous studies (for example, please see some references below, and references cited in these studies) investigating and comparing different degree-day snowmelt models and studies investigating calibration of conceptual hydrologic models (their snow part) to MODIS snow cover data. The introduction needs to clearly present the research done so far and to formulate what the novel scientific contribution of this study is. In my opinion, a comparison of existing degree-day models is not novel. Nor a general use of MODIS snow cover data in hydrological modelling. Still, I think the study presents some interesting approaches which can be turned into novel research objectives, such as how many and which MODIS images are needed for robust calibration of conceptual snowmelt *models.*

The introduction of the paper was certainly not complete, but this is not a review paper. In our understanding, the purpose of a non-review paper is to provide new ideas and not to give a complete picture of the state of the art. If the paper contains facts that are already known and not referenced, then this should be pointed out by the reviewers. Recently publications have extremely long review like introductions and number of cited papers increased significantly in the past years. In the opinion of the second author, this did not improve the quality of the publications.

However, with constructive feedback from the reviewer, have reformulated the Introduction section including a literature review of previous studies and the novelty of our approach in form of research objectives.

The structure of the document/story is not easy to understand, and the clarity of the presentation can be improved. If the study's main aim is to propose some novel approach/method, then I would suggest presenting it first and describing the study region and data later. This will allow the reader to understand the novelty and eventually to apply the general approach to other regions/models. I would also suggest presenting a general strategy at the beginning clearly. This will create a storyline and improves the clarity of the presentation. In its current form, there are many subsections and the order reads more like a summary of all technical works done but does not present clearly what the novel scientific contribution/research hypothesis is.

Thank you for this suggestion. We have tried to reconstruct the narrative in a more focused way. The general strategy, as suggested, has been modified. The subsections are removed wherever possible and the structure of the paper is streamlined to better exhibit the research idea and the results.

The study needs to be more focused on the novel contribution. I'm not sure how the interpolation and its cross-validation contributes to the novel scientific findings in the field of using satellite data form model calibration? Perhaps the crossvalidation can be presented only in a supplement. The more interesting point is to analyse which MODIS images are needed for robust model calibration. I do not understand why not use all available images, particularly for model validation? How sensitive are presented results to the selection of dates of MODIS images? There should be a more detailed analysis and evaluation for supporting the results and interpretations made. It is also not clear why not to use the concept of the HBV for simulation of snowmelt accumulation and melt. Why is it needed to separate the degree-day part and then link it back with the hydrological modules instead of using it together (i.e. to calibrate only the snow module first and then apply the complete model)?

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 have shortened this part of the paper and omitted from the Results section. Furthermore, we have added catchment based snowmelt models using the winter images for the entire time-period used in the study. In our opinion, the separation of the calibration and validation of the snow model from the hydrological model makes sense because of many reasons. Some of them are: (i) different hydrological models may use the same snow model (ii) an independently calibrated snow model reduces the uncertainty of the model parameters as no compensation of the model errors is possible through model parameters (reduced equifinality) (iii) snow models may be used individually for the estimation of available resources.

The discussion of the results is not comprehensive. It will be interesting to link the findings with similar studies calibrating the hydrologic models by using MODIS or comparing different variants of degree-day models.

Thank you for this suggestion. The discussion and concluding sections are extended with literature review in the revised manuscript.

I believe the manuscript presents an interesting topic and can be an interesting contribution, but it needs a very substantial revision and extension.

We would like to thank you for your encouraging words. We have done our best to improve the paper in its revised form.

Specific comments

Which MODIS version is applied?

Version V6 for both MODIS Terra (MOD10A1.006) and MODIS Aqua (MYD10A1.006) were used in the study. This was added in the revision in Line 115.

Kriging. Was the spatial correlation model (semi-variogram) fitted separately for each day?

Yes, the semi-variogram models were fitted for each day.

Radiation based model: how was the Linke coefficient estimated. Does it vary seasonally?

The Linke coefficient was set at a constant value of 3.0 (close to the annual mean for rural-city areas) for the model. Seasonal variation was not accounted for in the study but it could be an interesting addition. Instead, a diffusion factor (0.2 for clear sky to 0.8 for overcast conditions) was used to account for the diffusion.

Cross-validation of interpolation. Leave-one –out crossvalidation is typically used to camper different interpolation methods. Were the residuals smaller than obtained by some other interpolation method? How do the resulting maps compare with existing gridded (precipitation, air temperature) products provided by MeteoSwiss or DWD?

The goal of presenting the cross validation results was to give some information on input parameter uncertainty. The presented results suggest the Residual Kriging worked better than other methods for precipitation and External drift Kriging worked better for temperature data. However, the detail explanation of kriging and its results, due to the lesser relevance to the overall study, has been removed in the manuscript.

We have not compared the Kriged surfaces with the DWD or Meteoswiss gridded products yet.

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 have however, followed the suggestions of the reviewer and extensively rewritten an introduction with section with references pertaining to the use of Remote Sensing in modeling and prior approaches to improve the models in its revised form. We have also added references to each model formulation as suggested in the Methodology section.

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.

The sections are reformulated and mini-sections removed in the revised manuscript.

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.

Thank you for your feedback. The findings are presented in a better way in the revised document.

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. This has been well-documented in the revised manuscript for better understanding. Our results support the hypothesis that 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, as an extension to this research.

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 calibrating on discharge (which is still required to calibrate the other model parameters) is not clearly discussed.

A discussion on this has been added to the revised 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. However, we have re-calibrated the modified HBV with melt outputs from the snow-model and explained in detail in the results and discussion.

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. We have added 3 more catchments to the study to validate our findings at catchment levels.

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 your this critical comment. We have used the upstream catchment of Reuss at Seedorf instead, in the revised version. We apologise for the mistake. Furthermore, results from two more Swiss catchments, Aare at Brienzwiler and Thur at Andelfingen were also added to the revised paper.

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. We have done the revision in a reasonable amount of 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 modified 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 has been rephrased with the approach done by Bàrdossy et al. 2020. Please refer above for the last query.

Bárdossy, A.; Anwar, F.; Seidel, J. Hydrological Modelling in Data Sparse Environment: Inverse Modelling of a Historical Flood Event. Water 2020, 12, 3242. https://doi.org/10.3390/w12113242

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

The feedback is well-noted. However, for the clarity of the presentation, as many parameters are involved, we have for now kept it as it is.

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

This has been added in the Study area and data section.

- remove unnecessary digits in the tables

The digits were 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 October 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. The winter periods are defined clearly in the revision in Lines 312-313

-Any conceptual water-streamflow transformation model has to be calibrated with the water input. If you change the input, you have to re-calibrate 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 have re-calibrated the modified HBV in the revision with the melt from snow-models as the standalone inputs. The results clearly discuss the implications and reduction in uncertainty 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 have shortened this part of the paper.

- reference for Residual Kriging?

This has been added in the revised document in Line 149.

- 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 was clearly mentioned in the revised document in Lines 312-313.

- 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 in the first submission. However, we have added a correction factor to account for the combined gauge undercatch and vegetation interception for all the snow-melt models and the HBV. Please refer to Eq. 2a.

- 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 an average period from October to June for Switzerland to demarcate the 'snow' season on a more regional level.

- 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 has been omitted from the revision.