

Authors' response to interactive comment by Reviewer #3 María José Polo

Black text: Reviewer comment

Blue text: Authors' response

We thank the reviewer for her valuable comments and suggestions to improve our contribution. Below we reply to each of them and explain how we will incorporate them into the manuscript.

This work analyzes the performance of different snow routines based on the degree-day method in the framework of the HBV hydrological model. For this, runoff together with other snow-related variables are simulated in a large number of basins in Alpine areas in Central Europe and then compared to different sets of observations. The routines include different modifications for the snow routine components in HBV. Despite the significant variability found among cases, the results identified an exponential snowmelt function as the best modification in terms of model performance, followed by the adoption of a seasonal degree-day factor; other processes, like refreezing, added little benefit to the model pointing out that complexity itself is not an advantage without careful model design. The work addresses an interesting topic for areas where physical modelling approaches demand larger data sets than the available observations, and it is very clearly presented. Despite the conclusions cannot be directly extrapolated to other snow regions in the world, the number of study cases cover a large area in Central Europe, where snow processes condition the hydrological response in many rivers. I have some observations that can be assessed by the Authors to emphasize the applicability of the results and the scope of the study; some minor comments are also included.

1. The work includes all the different snow routines in the HBV model, and no other hydrological model is assessed. I suggest making it clear in the title that the assessment is done on the HBV performance, since "...for runoff modelling in mountainous areas in Central Europe", since it may lead to expect a wider scope of models. Additionally, some comments addressing whether the level of improvement or not obtained from each routine is affected by the model choice. At least, some reference to similar models should be included and some justification of what conclusions would be expected to be shared from simulations by other hydrological models.

Indeed, this study is focused on the HBV model as all the analyses were done using this specific model. However, we think of this study as having a wider scope than HBV, in that we propose a methodology to evaluate the impact of using different model structures for a large array of catchments in hydrological models that use the degree-day method to simulate snow processes. In this respect, also the related comment by Juraj Parajka is interesting. Actually, he rather asked for interpreting our results more broadly beyond the relevance for just the HBV model. He argued that this study might be interesting for other degree-day models, and asked to include some reference to the different implementations of this method in different hydrological models in the introduction. From this perspective, HBV is just the tool to show and evaluate this methodology. Nevertheless, we understand the concern of the reviewer, and we will expand the introduction and discussion sections to clarify which aspects of our study are specific to the HBV model and which are of broader relevance for other hydrological models that use the degree-day approach.

2. A second issue is related to the spatial resolution of the input data, and potential scale effects. Gridded weather data in the Swiss cases, 1-km² of gridded SWE, and 25-m cell size of the DEM, whereas point observations from stations and a 5-m DEM are used in the Czech catchments. Could you provide some assessment on these potential scale effects, and whether the source of weather data had an influence or not on the results? I also wonder whether using mean SWE values over each elevation zone, and point SWE measures, depending on the cases, could affect the results and comparison. Also, do you think that the results are scale-dependent of the cell size of the DEM used in the HBV model?

Regarding the DEM resolution, the cell size might have an impact on the results, but we argue that the proportions of the different elevation bands are represented correctly for both 5m and 25m resolution of the DEMs for most catchments in this study. This effect could become significant if the DEM would have a much coarser resolution (e.g. 500m) or if the catchments would be very small. In our case, we might only expect some minor effect in catchments with area less than about 10km² (which are only two of the 54 selected catchments). The effect of, for instance, the limited number of elevation bands (and the discontinuous and somewhat arbitrary choice of their elevation ranges) is probably much larger. Additionally, this factor may also be of importance for the snow model used to obtain the validation snow water equivalent data for the Swiss catchments, as topographical parameters such as slope and aspect need to be derived to correct for the influence of topography on snow distribution and redistribution.

Regarding the meteorological and SWE data, high-resolution data can become highly uncertain for individual points/grid cells, and these data should always be considered for somewhat larger areas. On the other hand, potentially high measurement errors and representativeness issues of the locality for the entire catchment/elevation band are also issues with observational data. We agree that the different approaches, i.e. catchment-wide aggregation of the gridded data product respective station data, might influence the results but its impact is hard to quantify. That being said, we would expect that the model performance variability resulting from individual model structures would be similar. We will mention these potential effects and their implications in the revised manuscript.

3. In the introduction, I miss some inclusions, like the importance of sublimation from the snow under certain conditions (not only in dry areas like we reported in Sierra Nevada-Spain, but also during the summer in the Alps and other regions, see Herrero and Polo, 2016), the existence of experimental catchments in the world devoted to snow processes research (see for example a recent Special Issue in Earth System Science Data on “Hydrometeorological data from mountain and alpine research catchments”), or the use of remote sensing sources to provide data to monitor snow-packs and snowmelt (many examples can be found, e.g. Dietz et al. 2012). Lines 55-60 should also address the limitations of degree-day approaches, and when they, although simple, are not an option.

We thank the reviewer for pointing out these aspects that certainly will enrich the introduction and help to put this study into a broader context of snow hydrology. Nonetheless, we already had considered some of the suggestions by the reviewer but had at the time decided to leave them out to avoid the introduction becoming overly long. Other points, such as the limitations of the degree-day approaches (e.g. snow towers, page 3) were already included in the manuscript, but maybe not with enough emphasis. We will revise the introduction considering the suggestions by the reviewer.

4. I am curious about the performance of each routine regarding the snow cover distribution. Did you check also their ability to capture this by testing against some satellite images? This is very interesting in terms of model performance to identify the sources of improvement or not.

We did consider using snow cover fraction as an evaluation metric for this study and performed some tests. However, in the end, we decided not to use it for different reasons. On one side, snow cover fraction does not provide a direct estimation of the amount of freshwater stored in the snow, which makes this parameter difficult to relate to the mass-balance approach of HBV. Additionally, cloud cover was an issue in the tests we performed. Besides, using this parameter could lead to large overestimations of snow water equivalent from, for instance, light snowfall events in late spring, when most of the catchment is no longer snow-covered but when there is still a significant storage of snow at high elevations, which would make the snow cover fraction jump up to 100% while the actual catchment-wide snow water equivalent would only have marginally increased. Finally, the scope of the study, including a large number of catchments and model alternatives, meant a large computational demand. We, therefore, made an effort to identify the most relevant metrics for evaluating the model for both snow processes and rainfall-runoff transformation. Considering additional metrics would certainly be very interesting and could add more value to the results but this is unfortunately beyond the scope of this study.

5. Since only four of the case studies were above 2000 m a.s.l. (only one above 2500 m), I think that some comment on how the results could change or not in higher elevation sites would shed light on their further applicability, especially in catchments where snowmelt is a higher fraction of runoff.

The reviewer raises an interesting question here. Indeed, only a handful of our catchments were at high elevations. There are few observations in high-elevation catchments and a lot of these catchments are influenced by glaciers. We took the decision to avoid glacierised catchments, as this would have required to increase the model complexity, and therefore the complexity of the analysis. This decision limited the number of suitable high-elevation catchments. It is difficult to speculate about the applicability of these results for high-elevation catchments, as they tend to be small, with steep topography and large glacierised areas, scarcely vegetated, and more exposed to extreme weather conditions such as strong wind gusts. Additionally, the applicability of the results would also be limited by a general limitation of degree-day methods, which leads to the occurrence of snow towers at high-elevations, where temperature hardly ever exceed the snowmelt threshold. We will include these considerations in the discussion.

6. I fully agree with selecting just some examples to conduct the presentation of results. I think, however, that including more than just one catchment, and year, would add value to your results. You could suggest another one from a lower altitudinal range, coming from the Swiss area, so that the impact of the spatial scale effects could, if needed, also be discussed. It would be very nice being able to see selected results from all the cases, I would suggest their inclusion as a supplement.

We agree with the reviewer in that including additional results, either more catchments or years, would improve the completeness of the manuscript and allow the reader to get more insights on the impacts of the different model modifications. Nevertheless, we feel that even including an additional catchment or year, would imply overly-extending the manuscript with additional figures and make the whole presentation of the results more cumbersome. Nevertheless, if the editor agrees we could include an appendix with figures similar to Figure 4 (including validation results, following a later comment by the reviewer) for all catchments. We include these figures at the end of the response for the reviewer to consider (caption only provided for Figure 1). This way, the interested reader will

be able to consult all the details while preserving a simple presentation of the results in the manuscript.

Other comments:

7. The gridded data of SWE in the Swiss cases were derived from a temperature-index model. Could this bias the performance of the routines?

The temperature-index (TI) approach, in which the snow model we used to derive snow water equivalent is based on, includes a time-varying threshold temperature (Slater and Clark, 2006) to differentiate between snowfall and rain, and allows for mixed precipitation within a transition temperature range. Using topographical parameters such as slope and aspect, the model corrects for the influence of topography on snow distribution and redistribution. The model follows the parameterization proposed in Helbig et. al (2015) to derive fractional snow-covered area. Despite these features, using a TI approach for both the rainfall-runoff model as well as for the snow model providing validation data might indeed lead to some bias. Nevertheless, it has to be taken into account that the snow model makes use of a 3-dimensional sequential data assimilation (DA). The DA itself includes two methods which are based on spatially correlated error statistics. For snow accumulation, an optimal interpolation approach uses the snow water equivalent station data to correct the simulated snowfall amounts. Regarding snowmelt, an ensemble Kalman filter updates snowmelt rates as well as liquid water content. Finally, the combination of both data assimilation approaches results in corrections of modelled snow water equivalent within all 1 by 1 km grid cells. Magnusson et al. (2014) investigate the performance in predicting snow water equivalent when using this DA approach and compare it to the TI model without DA. Based on 1033 samples from 45 stations, they show that using DA leads to improved snow water equivalent predictions.

8. Lines 259-260. Please, could you assess whether this decision could affect the results or not.

This decision might indeed have affected results, but the alternative would have caused a tremendous increase in parameter uncertainty and, thus, would have made the analyses almost impossible. In most of our catchments, elevation is the most important control on the spatial variation of snow processes, and this aspect is explicitly considered by using the elevation bands (using somewhat wider/narrower bands would likely have minor impacts on results, see Uhlenbrook et al., 1999). The implicit consideration of different vegetation types in one vegetation zone is frequently used in catchment modelling to avoid over-parameterisation.

9. Figure 4. Please, could you show also some validation results for this example case and year.

We understand that having some validation results would allow the reader to better assess the model performance as well as the modifications presented in this study. We will, therefore, modify Figure 4 to include validation results in addition to the calibration results.

10. Lines 410-412. Any comment on why these different behaviours are found?

Each individual modification of the snow routine adds between 1 and 2 additional parameters to the model. The design of HBV allows different parameters (even in different routines) to compensate for each other when calibrating the model. This issue is difficult to control for, especially when using automatic model calibration. Additionally, increasing the number of model parameters can also lead to over-parameterisation and equifinality issues. These different issues may lead to sub-optimal or physically inconsistent parameter sets that perform poorly when validating the model for an independent period. These potential issues lead us to be very careful in the model structures modifications we considered so as not to add too many additional parameters to the model.

11. Lines 425-427. Reading this, I would conclude that runoff data/simulations are somehow limiting the model performance's improvement (see also your comments in lines 482-484, and in lines 496-499). Additionally, this content should be reflected in conclusions (lines 565-567), to be more specific.

Good point. Yes, the evaluation against runoff data results in much smaller performance differences between the different model structures than the evaluation against snow water equivalent. This is to be expected as the ability of the model to simulate stream runoff is not only related to the structure of the snow routine but it is also affected by all other model routines that were not assessed in this study. We were aware of this issue but decided to perform the overall evaluation using these two objective functions based on two main considerations. First, if we want to evaluate changes on a particular routine of the model, we need to do it based on the output from the routine, not from the entire model, otherwise the noise from other routines of the model makes it impossible to attribute performance differences to any modification. That is why we used a metric based on snow water equivalent. Second, we were aware that HBV is not a perfect model and that it has issues with parameter compensation among others, and that the main application of the model is to simulate stream runoff. We, therefore, wanted to ensure that the modifications we introduced to the model were meaningful and produce acceptable results despite of its imperfect nature. These two considerations were equally important to us and that is why we evaluated the evaluations in this way, even at the cost of obtaining relatively modest results. We will include this in the revised conclusions.

12. I would suggest including some quantitative result in the conclusions, but I leave it up to the Authors.

We will consider this suggestion in the revision. Quantitative results are related to our particular set of catchments and the choice of using the HBV model. In contrast, the broader implications of our study might be more challenging to express in quantitative terms.

I hope that these comments help the Authors to address further their results and can contribute to the final version of the manuscript.

We thank the reviewer again for the helpful comments that will certainly improve the quality of our manuscript.

References

Dietz et al., 2012. International Journal of Remote Sensing a33(13): 4094-4134

<https://doi.org/10.1080/01431161.2011.640964>

Helbig, N., van Herwijnen, A., Magnusson, J., & Jonas, T. (2015). Fractional snow-covered area parameterization over complex topography. *Hydrology & Earth System Sciences*, 19(3).

Herrero and Polo, 2016. The Cryosphere, 10, 2981–2998, 2016 <https://doi.org/10.5194/tc-10-2981-2016>

Magnusson, J., Gustafsson, D., Hüsler, F., & Jonas, T. (2014). Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods. *Water resources research*, 50(10), 7816-7835.

Slater, A. G., & Clark, M. P. (2006). Snow data assimilation via an ensemble Kalman filter. *Journal of Hydrometeorology*, 7(3), 478-493.

Special Issue in Earth System Science Data on “Hydrometeorological data from mountain and alpine research catchments” https://www.earth-syst-sci-data.net/special_issue871.html

Uhlenbrook, S., Seibert, J. A. N., Leibundgut, C., & Rodhe, A. (1999). Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. *Hydrological Sciences Journal*, 44(5), 779-797.

Figures

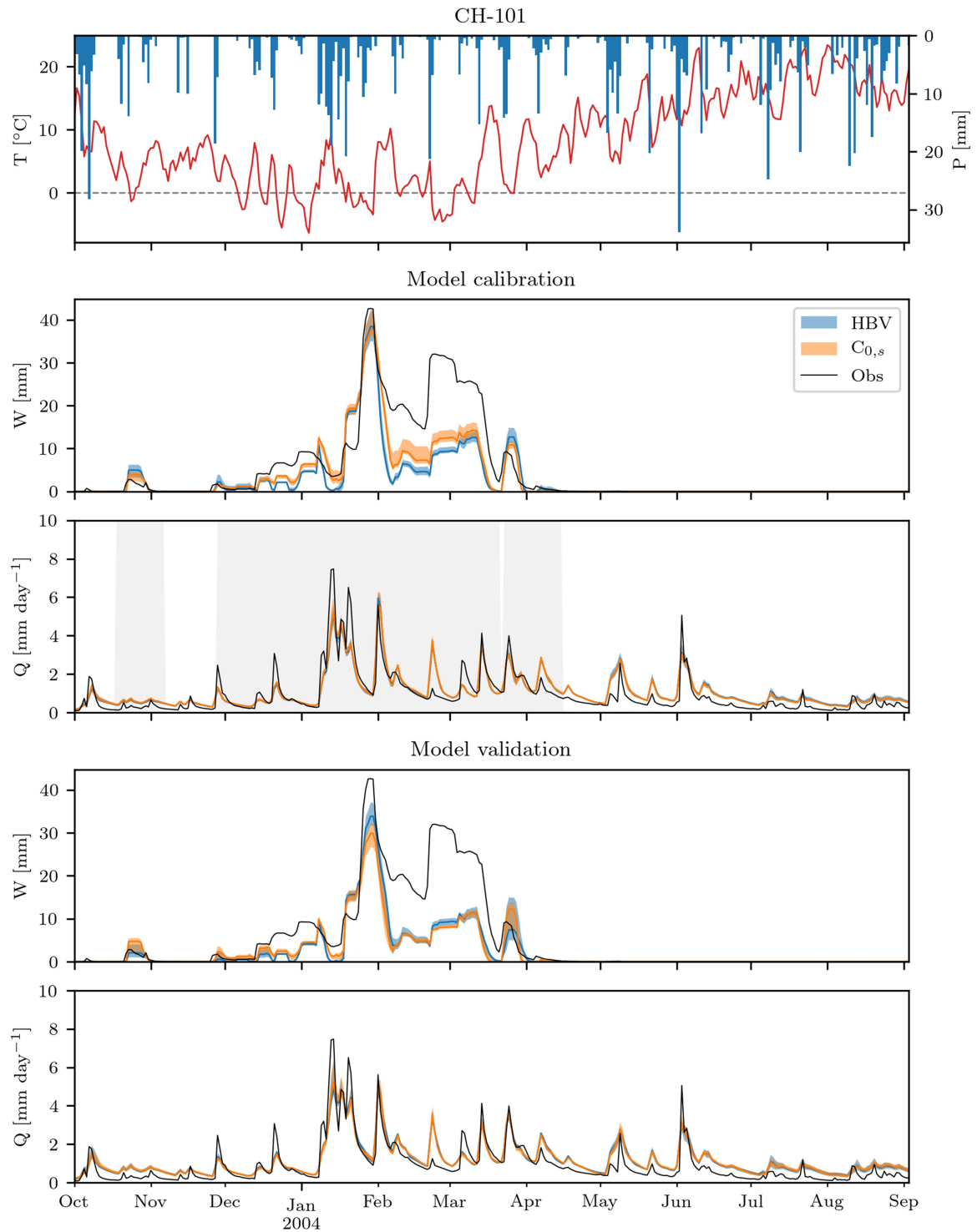


Figure 1 Time series (October 2003 – September 2004) for the catchment CH-101. Top: daily mean air temperature and total precipitation. Middle: model calibration results. Bottom: model validation results. The model calibration and validation are further subdivided into (top) catchment-average observed (grey line) and simulated snow water equivalent (HBV in blue and the model structure modification including a seasonally-varying degree-day factor, $C_{0,s}$ in orange), and (bottom) observed (grey line) and simulated stream runoff (HBV in blue and the model structure modification including a seasonal degree-day factor in orange). The grey field represents the period used when

calibrating the model against the logarithmic stream runoff. The uncertainty fields for model simulation cover the 10th – 90th percentiles range while the solid line represents the median value.

