Dear HESS Topical Editor Elena Toth,

Thank you for your efforts with our manuscript. We have now addressed the issues raised by the reviewers and included most of their comments and suggestions. Below we provide a point-by-point reply to each of the comments (blue text) as well as an explanation on how we included them in the text or, in the case where we did not include them, why we did not do so (red text). Among other changes, we adapted the title of the manuscript to better reflect the significance of this contribution. Additionally, we made general language improvements to the manuscript to improve its readability.

Thereafter we provide a marked-up version of the revised manuscript so that all the modifications from the previous version can be tracked. In this revised version we included an appendix containing a table of HBV model performance values for all catchments. Additionally, in a separate file we provide a supplement containing selected results (both for model calibration and validation) for each catchment.

With the aforementioned modifications and additions to the original manuscript we hope that this contribution meets the quality requirements to be published at Hydrology and Earth System Sciences.

Kind regards,

Marc Girons Lopez, Marc Vis, Michal Jenicek, Nena Griessinger and Jan Seibert

Authors' response to interactive comment by Reviewer #1 Juraj Parajka

We thank the reviewer for his valuable comments and suggestions that will help us improve our manuscript. Below we reply to each of these and explain how we incorporated them into the manuscript.

The study evaluates snow and runoff performance of 64 snow routine alternatives based on degree-day approach in large sample of catchments (54) located in Swiss and Czech Republic. The snow routine variants are coupled with HBV conceptual hydrologic model and model simulations are evaluated in terms of observed daily runoff and snow water equivalent observations/estimates. The results indicate that exponential snowmelt function with no refreezing and seasonally variable degree-day factors are the most reliable/robust/accurate variants for snowmelt runoff simulations in selected catchments. Overall, this is an interesting study which is worth to publish. The topic is relevant and within the scope of the journal. The study is clearly written and has a good structure. The analyses and interpretations are based on larger sample of catchments which allows to draw interpretations/conclusions that are relevant for large region of similar physiographic conditions in Central Europe.

Thanks for these kind words.

I have only few comments/notes which can be considered (in my opinion) to add/extend/improve clarity and generality of findings. These include:

1) Perhaps it will be possible to refer here in general to variants of degree-day snow approach, not strictly limits the analysis to HBV variants. The results can be used/implemented in degree-day routines of different hydrological models. In this study, the variants are coupled with HBV concept of rainfall-runoff transformation, but I believe, at least the evaluation of snow efficiency is relevant to general degree-day approach.

We also agree that the evaluation of the snow simulations is relevant to the degree-day approach in general, beyond its use in the HBV model. Indeed, what we present here is a methodology to analyse the impact of using different alternative model structures for a specific purpose (here, snow processes) on the performance of a rainfall-runoff model over a large sample of catchments. We, of course, related the proposed modifications to the HBV model, as this is our tool to conduct the analysis. Nevertheless, the alternative model structures that we explored in this study are not only "HBV variants", but at the same time also variants of the degree-day approach as used in other hydrological models as well.

We adapted the manuscript in general and more specifically the introduction to emphasise that our study is an evaluation of the degree-day approach in general which uses the HBV model as a tool to conduct the investigation.

2) When coupling the 64 snow routine variants with HBV model, there is another interesting question, which can be discussed and this is the robustness/uncertainty of other HBV model parameters. How consistent/different are the other HBV model parameters for different snow variants? Are, for example, field capacity or nonlinear runoff generation (beta) parameter values similar or compensating some effects of different snow routines?

The reviewer raises an important point here. In models such as the HBV model, model parameters can compensate each other, which makes the interpretation of any modifications rather challenging. While we did not include this in the manuscript, we performed a Monte Carlo sensitivity analysis on the HBV parameters. We provide figures resulting from these analyses for each catchment at the end of this comment (caption only provided for Figure 1). We found that, even if some of the variants (i.e. $T_{p,m}$, ΔP_e, or $C_{0,s}$) produce compensating effects on some parameters (e.g. PCALT, FC, LP or BETA), this effect was only observed for some of the catchments. Overall, parameter values and sensitivity tended to be fairly consistent across all the tested model variants for most of the catchments. We are aware of these potential compensatory effects between model parameters which can mask the real impact of different snow routine variants and, therefore, decided to base the evaluation of the analysis on the ability of the different model variants to reproduce snow water equivalent in addition to stream runoff (which is how hydrological models are traditionally evaluated).

In the revised version of the manuscript, we mentioned and discussed the sensitivity analysis both in the methods, results, and discussion sections (paragraphs starting at lines 329, 508, and 572) but we did not include any further figures in order to not make the manuscript even more complex.

3) It is not clear which part of the snow accumulation/melt phases are described/evaluated by selected snow objective function? For some practical applications, for example, it will be interesting to see the difference in maximum snow water equivalent between the routines, or to what extent the model over or underestimates snow cover duration? To what extent are these aspects covered in current snow efficiency evaluation? Does a good simulation mean well represented maximum SWE or snow cover duration? Perhaps there are some differences in such efficiency between the variants.

Our evaluation of the performance of the snow simulations provides an overall assessment of the snow processes in the selected catchments. As the reviewer correctly states, the evaluation could also be based on more specific aspects such as magnitude or timing of maximum annual snow accumulation. The problem with measures based on specific aspects is that a perfect fit with regard to one single measure does not ensure a good overall performance (similar to individual flow indices or signatures in the case of runoff simulations, see Vis et al., 2015). This implies that a number of measures would be needed to be used together with the challenge to decide on appropriate ways to combine the measures into a single overall performance measure.

While we agree that this could allow further assessments and might be valuable for future studies, we are afraid that such an additional analysis would be beyond the scope of this manuscript.

4) In our recent study (Sleziak et al., 2020) we found that there are quite significant differences in snow model performance (by using standard HBV degree-day approach) between lowland and alpine catchments in Austria. (Differences in terms of overestimation of snow cover in alpine and underestimation of snow cover in flatland catchments). Did you observe similar findings here?

Yes, in our study we also observe that model performance is generally lower for lowland catchments than for alpine ones. Our simulations also tended to underestimate snow water equivalent in lowland catchments. Regarding alpine catchments, we observed no clear pattern regarding over- or underestimation. Overall, it was quite frequent for the model to underestimate snow accumulation and delay the timing of the spring snowmelt season.

We appreciate the comment of the reviewer. We expanded the discussion of the manuscript and relate our results with the findings from the suggested article (Sleziak et al., 2020) (paragraph starting at line 547).

Specific comments

1) Abstract. Is the last sentence needed?

With this sentence we wanted to highlight some of the limitations of the results obtained in this study. However, as we also discuss these limitations in detail in the manuscript this "general disclaimer" might indeed not be needed.

We removed the last sentence from the abstract.

2) Introduction: It will be interesting to extend somewhat this section by referring to ways how can be/are degree-day routine parameters estimated in hydrological models.

We already do this to some extent in the methods section, where we present the different alternative model structures that we considered in this study, together with references. However, we understand that it would be relevant to refer to these different approaches and implementations in a general way in the introduction as well.

We expanded the introduction to give an overview of how different hydrological models use the degree-day method (paragraph starting at line 70) (see also response to major comment #1).

3) Data: How close are gridded snow water equivalent data to observations? Is there some bias related to the fact that this dataset is based on some type of degree-day model?

Using a temperature-index (TI) approach in both the runoff model as well as in 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 both 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. Concerning 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 improvements in predicting snow water equivalent.

We included a brief discussion on the possible impact of using a degree-day model for both the hydrological model and the estimation of snow water equivalent for Swiss catchments in the methods section (paragraph starting at line 244).

4) Runoff model efficiency. Why only Nash-Sutcliffe based on logarithmic transformed discharges? It will be interesting to see also the model performance in terms of snowmelt runoff peaks.

When we designed the study, we gave some thought on potential evaluation metrics, among which were the NSE, MARE or snow cover fraction. Nevertheless, since the computational demands for conducting this study were considerable (large array of catchments and model modifications), we decided to just use two objective functions (one for snow processes and another for rainfall-runoff transformation) that would be as relevant as possible. We agree that testing the model performance in terms of snowmelt runoff peaks would be very interesting indeed.

Looking at additional performance measures would be a valuable next step, but including this here would make the study overly complex due to the inclusion of too many aspects (see also answer to major comment #3).

5) P.15, l.355: Figure 3 or Figure 4?

This should indeed be Figure 4.

We corrected the error.

6) Figure 4. Will it be possible to show such case for a year in the validation period?

The intention behind presenting only the calibration results was to keep a simple story and walk the reader through the results by adding complexity stepwise. Nevertheless, we understand that having some detailed validation results might add valuable information to the reader as well.

We modified the figure to include validation results for the same year.

7) Results: Will it be possible to present runoff and snow model efficiencies for each catchment in the Supplement?

This study includes many different catchments and model variants, in addition to two periods for cross-validation analysis using two different metrics and presenting all these data in a meaningful way is not easy. However, since we only presented absolute results for one catchment in the manuscript, we agree with the reviewer that it would be a good idea to, at least, provide summarised results from all the catchments, periods, and evaluation metrics in an appendix.

In an appendix to the manuscript, we included a table showing median efficiency values for each catchment, and objective function, both for model calibration and validation for both periods for the default HBV model. Additionally, we included a supplement containing further figures similar to Figure 4 (including validation results, following the previous comment), for all catchments.

References

Sleziak et al. (2020) The effect of the snow weighting on the temporal stability of hydrologic model efficiency and parameters, Journal of Hydrology.

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.

Vis, M., Knight, R., Pool, S., Wolfe, W., & Seibert, J. (2015). Model calibration criteria for estimating ecological flow characteristics. *Water*, *7*(5), 2358-2381.

Authors' response to interactive comment by Reviewer #2 Giacomo Bertoldi

In attach some specific comments of Valentina Premier Ph.D. working with me.

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

The paper applies some modifications on the snow routine of the HBV model. Main results are that an increasing complexity does not lead to increasing performance. The most positively influencing modification is the use of an exponential snowmelt function and of a seasonally variable degree-day factor.

Some comments follow:

- Line 15-17: "However, [...] support tool" This sentence is not really clear to me. In general, I would restructure the abstract making clear from the beginning that the investigations are performed among snow routines based on temperature-index methods only.

With this sentence we wanted to point out that the implications of the decisions on which model structure to use for a given application are not always adequately addressed.

We rephrase this sentence to clarify this point. Additionally, we revised the abstract to clarify that the study builds on temperature-index methods.

- Line 34: "... often triggered by raising temperature". Is the main triggering source induced by air temperature or by incoming solar radiation, which is well represented by temperature?

Incoming solar radiation is indeed an important driver behind snowmelt, perhaps the most important one for open areas. This parameter is also strongly correlated with air temperature. Nevertheless, there are fluctuations in temperature that cannot be explained by incoming solar radiation alone, but by other processes such as lateral energy transfers, among others. For instance, snow also melts in locations with very little direct sunlight by the effect of temperature alone, such as under the canopy.

We clarified the text to make it more specific in respect to this, mentioning the important contribution of incoming solar radiation and its correlation with temperature (paragraph starting at line 32).

- Line 64-66: "Regarding the proportionality constant ..." Is the constant catchment defined? Are there studies which take into account of the spatial variability (e.g. different altitude, topography, etc?)

Yes, since we only use a single vegetation zone per catchment (see paragraph starting at line 117), the proportionality constant is catchment-defined. By defining different vegetation zones this parameter could take into account e.g. aspect, forested areas vs bare ground, etc. This would however come at the cost of having additional free parameters for calibration and make this study overly-complex. Other studies have indeed focused on the use of a spatially-variable proportionality constant (see e.g. He et al 2014).

We clarified this in the revised manuscript (paragraph starting at line 70).

- Line 66-67: ".. one for temperature and another for net radiation". Doesn't this belong to the hybrid methods?

Yes, the reviewer is correct.

We listed this approach under hybrid methods (paragraph starting at line 51).

- Line 115 Formula (3) Is T the daily average temperature? Some formulations take into account the cumulated temperature which exceeds the threshold, measured for example with 1 hour time step. Would these different formulation affect the results?

Yes, T refers to the daily average temperature, as it is common practice in degree-day approaches. Considering the approach mentioned by the reviewer is an interesting alternative, which might produce a somewhat increased simulated snowmelt, since the daily temperature pattern might allow for snowmelt during some hours, even if the daily average temperature is below the threshold for snowmelt.

This, however, is beyond the scope of our study, since it is limited to simulations at a daily resolution.

- Section 2.2.1 Has the formula (5) been evaluated by using the available temperature data for the studied catchment?

As mentioned in the manuscript, this equation is derived from the analysis of observational temperature data throughout the year from a large number of stations situated at different elevations (Rolland, 2003). We did not evaluate this equation here again since the temperature driving data we use in this study were either from a gridded data product based on the interpolation of station measurements (Switzerland) or single station measurements (Czechia). Based on these data, it is not possible to properly evaluate the equation. We did, however, check a sample year from a Swiss catchment for which we obtained the lapse rate from the gridded data product and fitted a constant and sinusoidal lapse rate parameter (Figure 1).

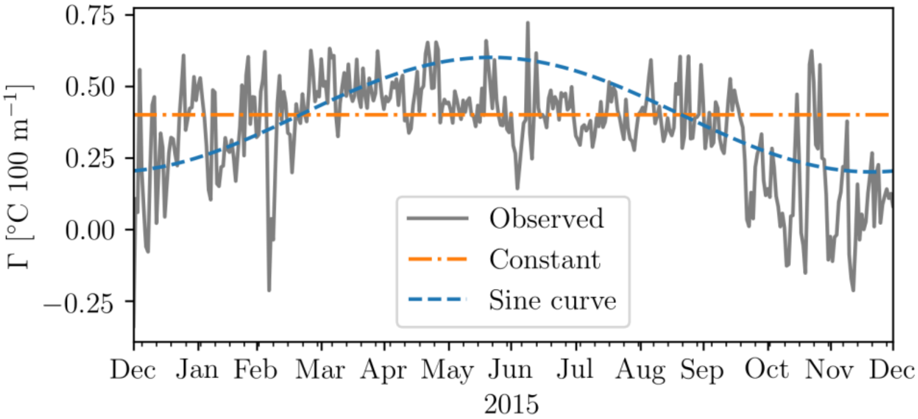


Figure 1. Comparison between the temperature lapse rate as described by a constant and sinusoidal parameters and the observed values from a gridded temperature data product.

- Paragraph 2.2.5 What is the threshold used in the model as the maximum liquid water content retained in the pores (maximum water retention capacity)?

We set the maximum liquid water content retained in the pores as a free parameter for calibration and restricted the range between 0 and 0.2 following Seibert (1999).

- Section Results. I would plot the performance vs size of the catchment and altitude (also for a fixed configuration, given the high number of variable components).

The reviewer makes a good suggestion. Actually, we expected to observe some relationship between these parameters and model performance and we tested this. We even tested other parameters such as yearly snowmelt contribution to runoff (we mention it briefly in the manuscript, lines 362-364). Nevertheless, we did not find any clear relationships for our case study.

References

He, Z. H., Parajka, J., Tian, F. Q. and Blöschl, G. (2014). Estimating degree-day factors from MODIS for snowmelt runoff modeling. *Hydrol. Earth Syst. Sci., 18,* 4773-8789.

Rolland, C. (2003). Spatial and seasonal variations of air temperature lapse rates in Alpine regions. *Journal of climate*, *16*(7), 1032-1046.

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and forest meteorology*, *98*, 279-293.

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

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 analizes 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. De-spite 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.

We expanded 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, and25-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 discussed these potential effects and their implications in the revised manuscript (paragraph starting at line 556).

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 revised the introduction considering the suggestions by the reviewer (paragraph starting at line 51).

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 above2500 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.

We commented on this in the revised version of the manuscript (paragraph starting at line 547).

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.

In an appendix to the manuscript, we included a table showing median efficiency values for each catchment, and objective function, both for model calibration and validation for both periods for the default HBV model. Additionally, we included a supplement containing further figures similar to Figure 4 (including validation results, following the previous comment), for all catchments.

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.

We included a brief discussion on the possible impact of using a degree-day model for both the hydrological model and the estimation of snow water equivalent for Swiss catchments in the methods section (paragraph starting at line 244).

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 modified 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 included this in the revised conclusions.

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

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.

We therefore decided not to include any quantitative results in the conclusions.

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.

Authors' response to interactive comment by Reviewer #4 Thomas Skaugen

We thank the reviewer for his 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.

General comments

This paper describes the testing of many alternative conceptual algorithms for snow modelling implemented in the Swedish HBV model. The suitability of the different algorithms has been assessed by split sample procedures for many catchments in Czechia and Switzerland. The paper is well written, well organized and the experimental setup seems, in principle, to be fine. However, the possible improvements of the tested alternative algorithms are extremely subtle and the authors recommend exponential snowmelt function and seasonally varying degree-day factor based on tiny improvements which have not, as far as I can see, been tested for their significance. I think the objective of the paper is good, it would be nice if we in objective ways could agree on improved concepts in snow modelling that when implemented would improve any model, but I am doubtful if the current methods are up for the task. The following issues need to be addressed in order to make paper suitable for publication.

Yes, the effects of the different modifications are small. Nevertheless, even if the average model performance improvement for the recommended model modifications are small on average, these modifications are significant for individual catchments, and do not lead to decreased model performance in any case. Minor changes had to be expected in general because we use a large sample of catchments in our study. Any improvements will, thus, tend to average out and look less impressive. We argue that while these improvements might indeed be small, our evaluation based on many catchments means that the findings are more robust than in many previous studies.

1) The paper misses a major investigation on equifinality issues (see papers of K. Beven and J. Kirchner on this topic). The HBV model itself has a lot of freedom, i.e. parameters to be calibrated, and most of the suggested algorithms for possible improved snow modelling add calibration parameters and hence to the problem of overparameterization. The point is that many of the suggested snow model modifications may have potential for being better at modelling snow, but the effect is impossible to isolate due to the overparameterization/equifinality. I have personal experience with trying to implement, what I thought was brilliant, ideas of improved snow modelling to the HBV model. They were all insignificant, and after a while I

realized that the compensating powers of all the parameters in HBV made it impossible to isolate and assess the effect of new algorithms (the frustration inspired the development of a new rainfall runoff model). The inclusion of the objective function for SWE is a step in the right direction, it narrows the freedom of the parameters, but probably not enough (you could try to also include Snow Covered Area, SCA). How many calibration parameters are there in the various model configurations? Are the numbers acceptable by any measure? Are their ranges physical at equifinality?

Indeed, equifinality is an issue in many hydrological models, and HBV is no exception to this. However, compared to many other models, the HBV model uses rather few parameters and parameter uncertainty is thus, smaller. The particular version used here, HBV-light, has been frequently used to address parameter uncertainty in the past years. So, while parameter uncertainty is an issue, we argue that we in the past have gathered quite some experience related to this issue. That being said, and while we did not include this in the manuscript, we performed a Monte Carlo sensitivity analysis on the HBV parameters. We found that, even if some of the variants (i.e. $T_{p,m}$, ΔP_e, or $C_{0,s}$) produce compensating effects on some parameters (e.g. PCALT, FC, LP or BETA), this effect was only observed for some of the catchments. Overall, parameter values and sensitivity tended to be fairly consistent across all the tested model variants for most of the catchments. As we previously mentioned, we took consideration of both potential equifinality and parameter compensation issues in our analysis.

We emphasised this in the revised manuscript. Furthermore, we mentioned and discussed the sensitivity analysis both in the methods, results, and discussion sections (paragraphs starting at lines 329, 508, and 572) but we did not include any further figures in order to not make the manuscript even more complex.

Most of the modifications that we tested in this study add only one extra parameter to the snow routine of the model (which consists of 5 parameters: degree-day factor, refreezing coefficient, threshold temperature, water holding capacity of the snowpack, and snowfall correction factor). Additionally, as the reviewer noticed, we assessed the impact of these modifications on the output of the snow routine (i.e., snow water equivalent) to avoid interactions from the other model routines and parameters. We also tested the use of other snow-related objective functions such as snow cover fraction. However, in the end, we decided not to use this measure because 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. Furthermore, using this objective function 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. Such events could 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 (i.e. Rw) and rainfall-runoff transformation (i.e. Rln(Q)). Regarding the decision to also evaluate the results respect to stream runoff, it was taken based on the common use of many hydrological models. We know that HBV is an imperfect hydrological model (as is the case for all models) based on certain assumptions that lead to issues such as parameter compensation. Even so, since models based on these assumptions will continue to be used in the foreseeable future mostly for runoff simulation, we wanted to ensure that the modifications we introduced to the model would produce acceptable results within the imperfect framework of the model. We agree with the reviewer in that better modelling approaches need to be found, but we also think that the available tools need to be evaluated and improved upon as well.

In the revised manuscript we clarified the choice of objective functions to perform the evaluation (paragraph starting at line 282).

Coming back to the number of additional calibration parameters for the various model configurations, this number varies between 1 and 3 parameters for single modifications. Several of the modifications that we selected were derived from observations, such as the seasonally-variable temperature lapse rate (Rolland, 2002) or the exponential precipitation phase partition (Magnusson, 2014). For the other parameters, we used constrained ranges (based on, for instance, Seibert (1999) for the default HBV structure) to ensure that parameter values do not become unrealistic. When different modifications are used simultaneously, the number may go up to 9 parameters. We regard this number as excessive and a clear over-parameterisation, which opposes the aim of preserving a simple structure with as few parameters as possible. However, we still included this variant in the evaluation for the sake of completeness.

We clarified this in the revised manuscript (paragraphs starting at line 329, 426, and 584).

2) I would desire a more stringent terminology. Words like “efficient” and “complex” have really lost their true meaning in the literature of hydrological modelling. Effective parameters really mean parameters that lump many processes or represents areal averages and has little to do with efficiency. A non-linear formulation of a process is not necessarily complex if the parameters are physical and measurable. To me, an over-parameterized model where, due to the compensating behavior among the parameters, the degree-day factor is suddenly correlated

to the parameter controlling the subsurface storage capacity is infinitely complex. Please consider rewriting the paragraph that starts at 525

We understand the concerns of the reviewer and agree with the need for a clear and concise terminology. Ideally, model efficiency relates to the definition the reviewer provides here. Nevertheless, as the reviewer also mentions, these terms can also be used to refer to other concepts such as models that provide acceptable results, even if for the wrong reasons. As explained in reply to the previous comment, in this study, we aim for both improving the quality of the processes conceptualisation in the model and ensuring that this improved conceptualisation works well with the imperfect nature of the model.

We revised the manuscript to ensure that the terminology we used was appropriate and concise. Additionally, rewrote the aforementioned paragraph to iron out inaccurate references to complexity and efficiency.

3) There are several paragraphs subjectively praising the HBV model for its ability to simulate hydrological behavior for various catchment types (p. 24, l.476, p.25, l 520-25, p.26, l563). “Hydrology” is a wide term and comprises more than runoff (and SWE admittedly), what about the subsurface, SCA, evapotranspiration etc. How come we are just presented result for one catchment?

The reviewer makes an important remark here to the need for objectively assessing the strengths and weaknesses of the selected models and design choices.

We revised the passages mentioned by the reviewer to ensure that the demands for objectivity were met adequately.

Regarding the term “hydrology”, we agree with the reviewer in that hydrology is a wide term and that there are many relevant variables and processes that often get overlooked in favour of – in most cases – stream runoff. We refer to the reply to major comment 1 for an explanation on the variables we used to assess the analysis presented in this contribution.

Similar to the previous point, we considered this comment when revising the manuscript and we replaced the generic references to “hydrology” by more concise terms.

We only presented results for a single catchment to not make the paper excessively long or complex. The analysis we present here includes a large array of catchments and alternative model variants, which make it impractical to present all the results in a detailed way without making the manuscript overly cumbersome.

In an appendix to the manuscript, we included a table showing median efficiency values for each catchment, and objective function, both for model calibration and validation for both periods for the default HBV model.

Additionally, we included a supplement containing further figures similar to Figure 4 (including validation results, following the previous comment), for all catchments.

Specific comments

P1, l.18, "popular" subjective

We agree with the reviewer that this is a subjective expression.

We rephrased the text to emphasise that this model has been (and still is) widely used in many different settings.

P1, l.27 "optimal degree of realism", rephrase

We rephrased the expression based on the responses to the major comments above.

P2, l.143-44 How can "the limitations of data availability–" "pose a challenge to properly monitoring". Rephrase....

With this sentence we wanted to express that monitoring hydrological processes (and more specifically snow processes in this case) is challenging with limited observations.

We rephrased the sentence for clarification.

P2, l.45-46 "Furthermore..." This sentence does not relate to anything above.

The idea we wanted to express with the final part of this paragraph was that, if having (limited) observations already makes it challenging to properly assess the current evolution of the hydrological processes (in this case snow processes), to be able to predict their future evolution with – obviously – no observations on the potential changes is even more complicated.

We rephrased this sentence to make it fit better in the paragraph.

P2, l.52 ..available at..

We corrected this mistake.

P2, l.53. ..in a distributed way.. Not always, see Skaugen et al., 2018 (Hydrology Research)

We admit that we over-generalised the identification of energy-based approaches with distributed hydrological models. We thank the reviewer for providing this reference.

We addressed this by explicitly mentioning lumped and semi-distributed models that are based on these approaches (paragraph starting at line 51).

P2, l.58 ..relevant for.. Aren't they relevant everywhere?

Indeed, using radiation data in addition to temperature data is relevant everywhere. Nevertheless, the benefits of using these approaches are most notable in the catchments described in the aforementioned sentence than in other types of catchments, where the impact is more modest. We did test such an approach for this set of catchments and found that the improvements were small while requiring an additional data source and calibration parameter.

P2, l.62 ..distribution function.. What is this?

We intended to say “bucket-type model” / “conceptual model”.

We corrected this error in the revision.

P3, l.83-84. ..and investigated whether.. See major comment above.

We refer to the reply to major comment 1 and 2 above.

We revised the text to ensure that the terminology we used is concise and relevant.

P4, l.98. “well established”, what does this mean? is it good or just old

We would argue that it means a bit of both. Indeed, the degree-day approach is an old conceptualisation of snowmelt processes, which has been in use for a long time already, and it has been evaluated, tested, and implemented in many different studies and models due to its low data requirements and explanatory power.

We rephrased this expression to include these nuances.

P5, l.125...to it... Refers to HBV or the individual components

It refers to the snow routine of HBV.

We reformulated this to avoid confusions.

P5, l.125 is precipitation lapse rate missing in the table? Could we have all calibration parameters in the table?

We intended this table to present the proposed modifications to the snow routine of HBV so, since we decided not to test any alternative to the precipitation lapse-rate, we did not include it in this table.

We added this component to the table.

Regarding the calibration parameters, we argue that including all calibration parameters is out of the scope of this table.

We added a sentence in Section 2.1 (in which the individual parameters are described) summarising the number of calibration parameters in the snow routine of the model. The reader will then be able to easily calculate the number of calibration parameters that are needed for each model variant.

P5, l.131. Heading, “Temperature and precipitation lapse rates”

We modified the section header (see also the previous comment).

P7, l.189. ..if somewhat.. How is it more realistic

It is more realistic than the one used in HBV since it does not have an abrupt transition (i.e. snowmelt being 0 up to a threshold and increasing linearly thereafter) in the change of snowmelt rate. Nevertheless, it does require the use of an additional parameter to control for the smoothness of the snowmelt transition.

P12, l.257. ..higher model complexity.. why more complex if you increase the temporal resolution

We explain this in the following sentences. We argue that to represent these processes at a sub-diurnal time step correctly, we would need to include additional parameters to control for processes that become relevant at these resolutions. Nevertheless, "complex" might not be the appropriate term here; "detailed" might be more suitable in this context (see also the reply to major comment 2 above).

We rephrased this sentence to be more concise in the terminology.

P12, l.257-58. "Other factors..", please elaborate

With this sentence, we meant that factors such as the transport time of meltwater from the snowpack to the stream become relevant for sub-daily time steps.

P12, l.266-67. Good, this fights the problem of over-parameterization. Could even include SCA

We refer to the reply to major comment 1 for a discussion on why we finally did not use snow cover fraction as a metric to evaluate this study.

P13, l.292. "efficiency", efficient how? Faster? gets more done? or just better?

In here, we referred to model performance when evaluating the model against each of the chosen objective functions.

As already mentioned in major comment 2 above, we revised the manuscript to ensure that the terminology is concise and relevant throughout the text.

P14, l.322..performance for...

We corrected the error.

P14, l.323-24. Can we accept improved performance for snow and decreased performance for runoff? I know that other authors have reported this, but is this not a clear indication of model structure flaw? Please elaborate, this is important

The reviewer raises an important point here, which is linked to the major comment above. Indeed, this kind of observations should make the hydrological community think about the complexity issues and limitations of the current generation of models and use this evidence to guide further research efforts that allow us to increase our understanding of these processes (including their connexions and feedback mechanisms) as well as to design and implement better (and usable) modelling strategies that avoid these issues.

Nevertheless, this issue is beyond the scope of this manuscript, which attempts to improve an existing, yet imperfect tool by exploring, testing, and evaluating the suitability of existing alternative structures.

P18, l.373.. “catchment dependent..” I do not understand this sentence

With this sentence, we wanted to express that there are some modifications which have a clear and consistent impact on most catchments (either negative or positive), while other modifications produce either positive or negative impacts depending on the catchment.

We rephrased the sentence to make it easier to understand.

P19, Figure. What does the y-axis represent, I struggle with this figure

The Y-axis shows the rank spread of each modification across all the catchments in the study, and each column adds up to 100%. So, for a given modification it shows for which percentage of catchments it is the best model structure, the second-best model structure, and so on. So, for instance, for the top left subplot, using a seasonal degree-day factor is the best alternative (among all the single modifications + default HBV structure) for ~80% of the tested catchments, the second-best alternative for ~10% of the catchments, and so on.

We clarified this in the text.

P20, l.404-407. This paragraph is very complex, can you please explain better

This paragraph is intended to clarify why, in the case of introducing 5 modifications to the snow routine (i.e. modifying each of the snow routine components that we evaluated), there are only three possible alternatives. The only available alternative representations (see also Table 1) are used for the lapse rate (i.e. Γ_s), the threshold temperature (i.e. $T_{P,M}$), the degree-day factor (i.e. $C_{0,s}$), and snowmelt and refreezing (i.e. M_e), in combination with one of the three alternative representations for the precipitation phase partition (i.e. ΔP_l, ΔP_s or ΔP_e).

We decided to remove this paragraph to avoid confusions.

P20, l.408 64 or 63 (see Table above)

We considered 64 different model structures, including the default HBV structure. In Table 3 we only showed modifications to the default HBV structure. Nevertheless, we understand that this can lead to confusions.

We changed the table accordingly.

P23, l.457 ..are dominant.. meaning strong or better?

In here, by “dominant” we meant the modifications to single components that appear most frequently in the top-ranking model variants.

We rephrased this passage to make it more concise.

P24, l.475. The first sentence is meaningless. Of course it is difficult to improve hydrological models, the processes are complex. The reason why it is difficult in the case of HBV could be due to the over-parameterization, not because it has been widely used with acceptable results.

This sentence was meant as an introduction to the discussion so, taking into account the major comments by the reviewer, we will modify just to state that we observed that it is difficult to improve hydrological models like HBV.

We included some comments in the discussion section on why such models are difficult to improve.

P24, l.487. .. runoff is modulated.. Rephrase

The intention with this sentence was to stress that, as already pointed out before, the final model output is the result of the interaction between the different routines of the model. This, as the reviewer points out in another comment, may be related to compensating effects between parameters but also to a loss of the signal from any modifications made on the snow routine. It is, therefore, to be expected that efficiency changes are minor when evaluating the model based on this variable.

We rephrased the sentence taking the previous discussion into account.

P25, l.533...even if model complexity...in a sensible way.. The sentence is strange

The intention with this sentence was to point out that, if there were enough data available and knowledge about the processes that to be simulated, then it would be justified to add more complexity (here understood as a more detailed description of the processes) to the model.

We rephrased the sentence to make it clearer.

P26, l.551 different settings.. please be more specific.

By different settings, we were referring to the geological, geographic, climatological, and hydrological characteristics that define the hydrological behaviour of a given catchment.

We made this sentence more specific.

P26, l.563. Unsubstantiated, we have only seen the result for one catchment.

We understand the concern of the reviewer regarding drawing general conclusions about the whole study when we only presented the details of a single catchment in the manuscript. We took this decision to facilitate the storyline of the manuscript and not overwhelm the reader with endless results.

We refer to major comment #3 for the relevant modifications concerning this comment that we included in the revised manuscript. We hope that the aforementioned changes provide enough evidence to support this conclusion.

P26, l.565. How to proceed with this “better approach”, how to do it in practice?

In this conclusion point, we state that carefully assessing which objective, the necessary level of detail (see previous comment on complexity vs detail), and data availability to each case is a better approach than just picking whichever model we are familiar with or have a preference for. Obviously, this is easier said than done but, in this contribution, we aim to provide a methodology to do such an assessment over a large sample of catchments and model structure variants for a specific purpose.

We clarified this point in the revised manuscript.

References

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.

Rolland, C. (2003). Spatial and seasonal variations of air temperature lapse rates in Alpine regions. *Journal of climate*, *16*(7), 1032-1046.

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and forest meteorology*, *98*, 279-293.

Skaugen, T., H. Luijting, T. Saloranta, D. Vikhamar-Schuler and K. Müller,2018. In search of operational snow model structures for the future - comparing four snow models for 17 catchments in Norway. Hydrology Research, 49.6, https://doi.org/10.2166/nh.2018.198

Complexity and performance Assessing the degree of detail of temperature-based snow routines for runoff modelling in mountainous areas in Central Europe

Marc Girons Lopez[1, 2], Marc J. P. Vis[1], Michal Jenicek[3], Nena Griessinger[4], Jan Seibert[1, 5]

[1]Department of Geography, University of Zurich, Zurich, CH-8006, Switzerland
[2]Swedish Meteorological and Hydrological Institute, Norrköping, SE-60176, Sweden
[3]Department of Physical Geography and Geoecology, Charles University, Prague, CZ-128431636, Czechia
[4]WSL Institute for Snow and Avalanche Research SLF, Davos, CH-7260, Switzerland
[5]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, SE-75007 Sweden

Correspondence to: Marc Girons Lopez (marc.girons@smhi.se)

Abstract. Snow processes are a key component of the water cycle in mountainous areas as well as in many areas of the mid- and high latitudes of the Earth. The complexity of these processes, coupled with the limited data available on them, has led to the development of different modelling approaches to aimed at improveing their our understanding of these processes and supporting management decision-making and management practices. Physically-based approaches, such as the energy balance method, provide the best representation of these snow processes but at the expense of high data requirements. Data limitations in data availability, in most many situations, constrain their applicability use of these methods in favour of more simple straightforward approaches. Indeed, Tthe comparatively simple temperature-index method is has become the most widely-used modelling approach of thefor representing snowpack processes in rainfall-runofffor hydrological modelling, with many different variants of this method implemented in differentacross many models. HoweverNevertheless, in many cases, the decisions on the most suitable complexity degree of detail of these conceptualisations the model are in many cases not adequately assessed for a given model structure, application, or decision-making support tool.

In this study, we assessed the model structure choicessuitability of a number of formulations of different components of the simple temperature-index method for rainfall-runoff modelling in mountainous areas of Central Europe by using the HBV bucket-type model. of the HBV model, a popular semi-distributed, bucket-type hydrological model, for its application in mountainous areas in Central Europe. To this end, we reviewed the most widely-used choices formulations toof different components of the temperature-based snow routines fromin different hydrological rainfall-runoff models and proposed a series of modifications to the default structure of HBVthe HBV model. We constrained narrowed the choice of modifications alternative formulations to those that provide a simple conceptualisation of the described processes in order to constrain parameter and model uncertaintyare aligned with HBV's modelling approach of keeping processes as simple as possible to constrain model complexity. We analysed a total of 64 alternative snow routine structures over 54 catchments using a split-sample test. We found that using (a)Overall, the most valuable modifications to the standard structure of the HBV snow routine were (a) using an exponential snowmelt function coupled with no refreezing, instead of a linear function for both processes and (b) computing melt rates with a seasonally-variable degree-day factor instead of a constant one were, overall, the most

~~valuable modifications to the model~~. ~~Additionally, we found~~Our results also demonstrated that increasing the degree of detail of the temperature-based snow routines in rainfall-runoff models ~~complexity does~~ did not necessarily lead to an improved model performance per se. Instead, ~~we found that a thorough~~performing an analysis ~~of the different~~on which processes are to be included, ~~in the model~~ and to which degree of detail, ~~their optimal degree of realism~~ for a given model and application is a ~~preferable alternative~~better approach to obtain more reliable and robust results. ~~While the results may not be transferrable to other modelling purposes or geographical domains, the methodology presented here may be used to assess the suitability of model design choices.~~

1 Introduction

Snow is an essential aspect of the ~~seasonal~~ annual hydrological variations in Alpine areas as well as in many other regions of the mid and high latitudes of the Earth. Unlike rainfall, which contributes directly to the groundwater recharge and stream runoff, snowfall accumulates on the ground creating a temporary freshwater reservoir. This accumulated water is then gradually released through melting when the necessary energy for melt is available, contributing to runoff.~~, often triggered by~~ Incoming solar radiation is the major control of the variability of the available energy whilst air~~raising~~ temperature is a good proxy for the variation of the available energy and, thus, snowmelt (Sicart et al., 2008)~~s, and ultimately contributes to runoff~~. The snow accumulated on the ground (i.e., snowpack) is not only crucial for ecological reasons (Hannah et al., 2007), but also for many human activities such as hydropower, agriculture, or tourism (Barnett et al., 2005). At the same time, snow processes can also lead to risks for society. For instance, the accumulation of snow on steep slopes may~~, under the right conditions,~~ cause avalanches (Schweizer et al., 2003), and -the sudden melt of large amounts of snow~~, such as~~ during rain-on-snow events (Sui and Koehler, 2001) or after a rapid increase~~s~~ of air temperature, may lead to widespread flooding ~~either directly~~ (Merz and Blöschl, 2003; Rico et al., 2008)~~ or indirectly (e.g. dam failure accidents) (Rico et al., 2008)~~.

Society´s dependence on the freshwater stored in the snowpack and its vulnerability to its associated risks raises the need to understand its dynamics and evolution (Fang et al., 2014; Jamieson and Stethem, 2002). Nevertheless, ~~even if~~while knowledge on snow hydrology has broadly advanced over the last decades with, for instance, the establishment of experimental catchments devoted to snow processes research (Pomeroy and Marks, 2020), or the use of remote sensing data for snowmelt monitoring (Dietz et al., 2012), ~~the limitations of data availability on catchment hydrology~~limited observations in most locations still ~~— and especially on snow processes —~~pose a challenge to properly ~~monitoring~~ quantifying snow~~these~~ processes ~~as well as~~and implementing adequate management policies and practices. ~~Furthermore~~In addition to present-day limitations, the evolution of snow water resources in the future, which cannot be estimated through direct observations, ~~but~~ is also essential in the context of global climate change (Berghuijs et al., 2014; Jenicek and Ledvinka, 2020).

~~Consequently, d~~Different modelling strategies have been developed to overcome the data limitations and to study the evolution of the snowpack and its impact on water resources. The most common modelling approaches are based either on the physically-based energy budget model or on the temperature-index ~~distribution function method~~approach (Verdhen et al., 2014). While energy budget models are the most accurate alternative to represent snowpack processes, in order to be reliable they usually require data that are often not available in at conventional meteorological stations (Avanzi et al., 2016). These models attempt to estimate the snow contribution to runoff, generally in a distributed way, by solving the energy balance of the snowpack, which requires detailed data on topography, temperature, wind speed and direction, cloud cover fraction, snow density, etc. Some efforts have also been done to implement such approaches at sub-catchment or even catchment scales, thus requiring less driving data (Skaugen et al., 2018). Conversely, tTemperature-based models (also known as temperature-index or degree-day methods), in contrast, are based on the assumption that the temporal variability of incoming solar radiation is well represented by the variations of air temperature (Ohmura, 2001; Sicart et al., 2008) and tend to have low, and thus easy to meet data requirements and computational demands, and offer a satisfactory balance between simplicity and realismperformance, which makes them successful in many different contexts and applications, even in cases with limited data availability (Hock, 2003). Nevertheless, the assumption that incoming solar radiation is well represented by air temperature does not always hold, such as in high elevation catchments where temperature seldom raises above the freezing point (Gabbi et al., 2014; Pellicciotti et al., 2005), or in conditions in which sublimation from the snowpack becomes a significant process (Herrero and Polo, 2016). Such issues led to the development of extended formulations including additional variables such as wind speed or relative humidity to improve the snowmelt estimation (Zuzel and Cox, 1975) or even Hhybrid methods combining energy-based and temperature-based approaches have also been developed, such as including the inclusion of a radiation component in the temperature-based models (Hock, 1999; Kane et al., 1997). These approaches are especially relevant for high elevation, often glacierised areas in which temperature seldom gets above the freezing point (Gabbi et al., 2014; Pellicciotti et al., 2005).

The tSimple temperature-index models define – also referred to as degree-day method – is based on the observation that the rate of snowmelt ias being proportional to the temperature above the freezing point per unit time through a proportionality constant commonly named degree-day factor (Collins, 1934; Martinec, 1960). Many distribution functionconceptual hydrological rainfall-runoff models use variations of this method to simulate snowpack processes. For instance, while many models use a simple formulation including a constant degree-day factor both in time and space (Valéry et al., 2014), others include a monthly or seasonally variable parameter (Hottelet et al., 1994; Quick and Pipes, 1977) or even a spatially variable degree-day factor that takes, amongst others, differences in slope, aspect, or vegetation cover into account (He et al., 2014). AdditionallyFor instance, while some models use the freezing point (i.e. 0°C) as the reference threshold temperature for the onset of snowmelt (Walter et al., 2005), others include a calibrated threshold temperature parameter (Viviroli et al., 2007) to allow for spatial variations on this process. Furthermore, some models disregard some of the processes, such as refreezing, as their magnitude tends to be negligible with respect to snowmelt (Magnusson et al., 2014). Regarding the proportionality constant, some applications define it as time-independent (Valéry et al., 2014), while others establish it as being seasonally variable (Hottelet et al., 1994). Moreover, connecting with energy budget models, some applications use two proportionality constants: one for temperature and another for net radiation (Kane et al., 1997). Other components of the snow routine may also be conceptualised with different degrees of detail. A good example is the formulation of the precipitation phase partition between rain and snow. While some models set a sharp threshold for this transition, others use a gradual transition where rain and snow may occur at the same time, using different model formulations and, in some cases, also additional data such as relative humidity (Matsuo and Sasyo, 1981) to define this transition. ~~Nevertheless~~In general, however, ~~regardless of the preferred approach,~~ the inherent simplifications made in semi-distributed temperature-index models leave out some critical aspects of the snowpack processes that may be significant in some circumstances. For instance, the disregarding of lateral transport processes in many models may lead to the development of unreasonable accumulations of snow over long periods ~~of time~~ (i.e., snow towers) in high mountainous areas (Freudiger et al., 2017; Frey and Holzmann, 2015).

Overall, the degree of detail in which the different snow processes are formulated in different models differs greatly, and depends to a great extent on the~~The different implementations of the temperature-index method may relate to differences in~~ model philosophy and preferences, purpose~~scope~~, application, desired resolution, or available data and computing power, among others. Nevertheless, these choices are not always adequately ~~tested~~ taken into account when using a specific model for a different application or purpose to what it was originally developed for ~~to ensure that they provide the best possible alternative for a given model design and application~~ (Harpold et al., 2017),~~. For instance~~ may lead~~, some studies have found~~ to models with ~~that~~ a more ~~realistic~~ detailed representation of hydrological processes performing worse than~~does not always translate into improved model performance~~ comparatively more simplistic models for a specific purpose (Orth et al., 2015). So, models (or the relevant model routines) should always be tested beforehand to ensure that the assumptions and formulations used are adequate and robust for the intended application (Günther et al., 2019). For a long time, however, limitations in computing power hindered the systematic testing of different ~~alternative~~ model structures over a large number of catchments. ~~I~~In recent years, however, ~~however,~~ the increase in computing power has made these tests not only feasible but also desirable ~~in order to ensure that model structures are adequate and robust for their intended applications (Günther et al., 2019)~~.

In this study, we present a methodology to evaluate the design choices of a rainfall-runoff model with a simple temperature-based snow routine for its application over a large number of catchments. More specifically, we aim to evaluate~~d~~ the ~~design choices made in~~suitability of the snow routine of the HBV ~~hydrological~~ rainfall-runoff model (Bergström, 1995), a typical bucket-type model ~~with a temperature-based snow routine~~, for its application in mountainous catchments in ~~c~~Central~~-~~ Europe~~an catchments~~. Taking the existing model structure as a reference, ~~W~~we implemented and tested a number of model structure modifications ~~to the snow routine of the model~~ based on common ~~practices~~formulations of snow processes in ~~similar~~ other ~~hydrological~~ rainfall-runoff models with simple temperature-based snow routines~~, and investigated whether it is possible~~ to ~~identify model structure alternatives which~~ assess the most suitable model structure for the intended application. That is, model structures which generally result in improved model performance for representing both snow processes and stream runoff in the area of interest~~. With this, we aimed to provide a basis to decide on useful modifications to the model~~ while avoiding adding unnecessary ~~complexity~~ elements and ~~additional~~ parameters that would result in increased model uncertainty and equifinality issues (Beven, 2008). To ensure that the results are representative, we explored different levels of added ~~complexity~~detail, from ~~single~~ modifications to single components of the snow routine to combinations of ~~multiple~~ modifications to multiple components~~,~~ on~~to~~ a large dataset of catchments covering a wide range of geographical, climatological, and hydrological conditions of the area of interest.

2 Materials and ~~M~~methods

The HBV model is a bucket-type rainfall-runoff model~~,~~ with a number of ~~boxes (~~routines~~)~~ ~~including~~representing the main components of the terrestrial ~~phase~~ part of the water cycle, i.e., snow routine, soil routine, groundwater (response) routine, and routing function. In this study, we focused solely on the snow routine of the model. We used the HBV-light software, which follows the general structure of other implementations of the HBV model and includes some additional functionalities such as Monte Carlo runs and a genetic algorithm for automated optimisation (Seibert and Vis, 2012). Henceforth we use the term 'HBV model' when referring to our simulations using the default HBV-light software.~~. With 'HBV model' we mean the 'standard HBV', i.e.,~~That is, the HBV model with the snow routine as described in Lindström et al. (1997) and~~or~~ Seibert & Vis (2012).

2.1 HBV's ~~S~~snow ~~R~~routine

The snow routine of the HBV model is based on ~~well-established~~widely-used and well-tested conceptualisations of the relevant snow processes for ~~hydrological applications~~rainfall-runoff modelling. More specifically, i~~I~~t represents the main processes ~~regarding two aspects of snow hydrology:~~ related to (i) the precipitation phase partition between ~~of the precipitation (~~snow and~~or~~ rain~~)~~, and (ii) the snow accumulation and subsequent melt and refreezing cycles of the snowpack.

Regarding the precipitation phase partition, HBV uses a threshold temperature parameter, T_T *[°C]*, above which all precipitation, P *[mm ~~/~~Δt^-1]*, is considered to fall as rain, P_R *[mm ~~/~~Δt^-1]* (Eq. 1). This threshold can be adjusted to account for local conditions. Below the threshold, all snow is considered to fall as snow, P_S *[mm ~~/~~Δt^-1]* (Eq. 2). The combined effect of snowfall undercatch and interception of snowfall by the vegetation is represented by a snowfall correction factor, C_{SF} *[-]*.

$$P_R = P, \; \&T > T_T \;, \tag{1}$$

$$P_S = P \cdot C_{SF}, \; \&T \leq T_T \;, \tag{2}$$

As previously mentioned, the HBV model uses a simple approach based on the temperature-index method to simulate the evolution of the snowpack. This way, snowmelt, M *[mm ~~/~~Δt^-1]*, is assumed to be proportional to the air temperature, T *[°C]*, above a predefined threshold temperature, T_T *[°C]*, through a proportionality coefficient, also called degree-day factor, C_0 *[mm ~~/~~Δt^-1 °C^-1]* (Eq. 3). ~~The physical motivation of this approach is that the energy available for snowmelt is generally proportional to the air temperature (Ohmura, 2001).~~ The model allows for a certain volume of melted water to remain within the snowpack, given as a fraction of the corresponding snow water equivalent of the snowpack, C_{WH} *[-]*. Finally, ~~the~~ refreezing of melted water, F *[mm ~~/~~Δt^-1]*~~,~~ takes place when the air temperature is below T_T, and its magnitude is modulated through an additional proportionality parameter, C_F *[-]* (Eq. 4).

$$M = C_0(T - T_T) \;, \tag{3}$$

$$F = C_F \cdot C_0(T_T - T) \;, \tag{4}$$

Overall, the snow routine of the HBV model contains five calibration parameters. HBV allows for a limited representation of catchment characteristics through the specification of different elevation and vegetation zones. This way, the parameters controlling the different processes included in the snow routine can be modified for individual vegetation zones. The combination of elevation and vegetation zones (also known as Elevation Vegetation Units, EVUs) is the equivalent of the Hydrologic Response Units (HRUs) used in other ~~distribution function~~conceptual models (Flügel, 1995). Both precipitation and temperature are corrected for elevation using respective~~two~~ lapse-rate parameters~~ for the precipitation and temperature lapse rate~~.

2.2 Proposed ~~M~~modifications to individual components of the ~~S~~snow ~~R~~routine~~ Components~~

Here we review the ~~individual~~ different components of the snow routine structure of the HBV model as well as functions that are directly related to this routine~~it~~ (e.g. input data correction with elevation) and describe the proposed modifications to each component. Each of these ~~modifications~~ alternative structures requires one to three additional parameters (Table 1).

Table 1 Description of the proposed modifications to the snow routine of the HBV model. The default component structures of the HBV model are marked with a * symbol. The components marked with a † are not formally part of the snow routine but were included in the analysis due to their significant impact on it.

Snow routine component	Structure	Abbreviation	Number of additional ~~P~~parameters
Precipitation lapse rate†	Constant*	-	1
Temperature lapse rate†	Constant*	Γ_c	1
	Seasonally-variable	Γ_s	2
Precipitation phase partition	Abrupt transition*	ΔP_a	1
	Partition defined by a linear function	ΔP_l	2
	Partition defined by a sine function	ΔP_s	2
	Partition defined by an exponential function	ΔP_e	2
Threshold temperature	One threshold for both precipitation and snowmelt*	T_T	1
	Different thresholds for precipitation and snowmelt	$T_{P,M}$	2
Degree-day factor	Constant*	$C_{0,c}$	1
	Seasonally-variable	$C_{0,s}$	2
Snowmelt and refreezing	Linear snowmelt and refreezing magnitude increase with temperature*	M_l	3
	Exponential snowmelt magnitude increase with temperature. No refreezing.	M_e	3

2.2.1 Temperature and precipitation ~~L~~lapse ~~R~~rates

When different elevation zones are used, the temperature for each zone is generally computed from some catchment-average value and a lapse rate parameter. In HBV, ~~usually~~ a constant temperature lapse rate is usually used. Alternatively, if the available data allows, it is also possible to provide an estimation of the daily temperature lapse rate. However, if no data on the altitude dependence of temperature is available, setting a constant value throughout the year might be an oversimplification. Indeed, in an experimental study on several locations across the Alps, Rolland (2002) found that the seasonal variability of the temperature lapse rate follows approximately a sine curve with a minimum a~~t~~round the winter solstice~~.~~. ~~Following these results~~Following these findings, we implemented a seasonally variable computation of the temperature lapse rate using a sine function (Eq. 5). This way, the temperature lapse rate for a given day of the year, Γ_n *[°C ~~/~~100 m^{-1}]* (where *n* is the day of the year, a sequential day number starting with ~~day~~ 1 on the 1st of January), depends on two parameters, namely the annual temperature lapse rate average, Γ_0 *[°C ~~/~~100 m^{-1}]*, and amplitude, Γ_i *[°C ~~/~~100 m^{-1}]*.

$$\Gamma_n = \Gamma_0 + \frac{1}{2}\Gamma_i \sin\frac{2\pi(n-81)}{365}\,, \tag{5}$$

Precipitation lapse rates ~~could~~ cannot be related to ~~a~~ seasonal or other types of systematic variations as they are strongly dependent on the synoptic meteorological conditions and therefore highly variable. ~~Therefore~~Consequently, we decided to keep the default approach in the HBV model which consists in~~of~~ calibrating the model using a~~n~~ ~~average~~ constant precipitation lapse rate parameter.

2.2.2 Precipitation ~~P~~phase ~~P~~partition

The determination of the precipitation phase is a crucial step as it controls whether water accumulates in the snowpack or contributes directly to recharge and runoff. In the HBV model~~,~~ the distinction between rainfall and snowfall is based on the assumption that precipitation falls either as rain or as snow, depending on a threshold temperature parameter. However, in reality, this transition is less sharp, as ~~there are mixed events with~~ both rain and snow can coincide (Dai, 2008; Magnusson et al., 2014; Sims and Liu, 2015). Additionally,~~and,~~ depending on other factors such as humidity and atmosphere stratification, the shift from rain to snow can occur at different temperatures ~~(Dai, 2008; Magnusson et al., 2014; Sims and Liu, 2015)~~. Therefore, the single threshold temperature may not adequately represent the snow accumulation, especially in areas or periods with temperatures close to zero degrees Celsius. Different ~~approaches~~ formulations have been ~~suggested~~ proposed to describe the snow fraction of precipitation, *S [-]*, as a function of temperature (Froidurot et al., 2014; Magnusson et al., 2014; Viviroli et al., 2007). In this study, we considered three different ~~conceptualisations~~ formulations to calculate ~~of~~ the snowfall fraction of precipitation~~,~~ (Eq. 6 ~~–~~and 7, respectively): (i) a linear function (Eq. 8), (ii) a sine function (Eq. 9), and (iii) an exponential function (Eq. 10). Both the T_A *[°C]* and M_P *[°C]* parameters control the range of temperatures for mixed precipitation.

$$P_S = P \cdot S \cdot C_{SF}\,, \tag{6}$$

$$P_R = P \cdot (1 - S)\,, \tag{7}$$

$$S = \begin{cases} 1, \ T \leq T_T - \frac{T_A}{2} \\ \frac{1}{2} + \frac{T_T - T}{T_A}, \ T_T - \frac{T_A}{2} < T \leq T_T + \frac{T_A}{2}, \\ 0, \ T > T_T + \frac{T_A}{2} \end{cases} \tag{8}$$

$$S = \begin{cases} 1, \ T \leq T_T - \frac{T_A}{2} \\ \frac{1}{2} - \frac{1}{2}\sin\left(\pi \frac{T_T - T}{T_A}\right), \ T_T - \frac{T_A}{2} < T \leq T_T + \frac{T_A}{2}, \\ 0, \ T > T_T + \frac{T_A}{2} \end{cases} \tag{9}$$

$$S = \frac{1}{1 + e^{\frac{T - T_T}{M_P}}}, \tag{10}$$

2.2.3 Snowmelt ~~T~~threshold ~~T~~temperature

In addition to determining the precipitation phase, a temperature threshold parameter is also needed to determine ~~when~~ the onset of snowmelt ~~starts~~. The most straightforward approach, ~~as~~ used in the HBV model, is to use the same threshold temperature parameter for both snowfall and snowmelt. However, as these two transitions are related to different processes happening at different environmental conditions, a single parameter might not adequately describe both transitions. A more realistic approach would be to consider two separate parameters for these processes: a threshold temperature parameter for precipitation phase ~~differentiation~~partitioning, T_P *[°C]*, and another one for snowmelt and refreezing processes, T_M *[°C]* (Debele et al., 2010)~~. We implemented this modification using one additional parameter~~.

2.2.4 Degree-day factor

The degree-day factor is an empirical factor that relates the rate of snowmelt to air temperature (Ohmura, 2001). In the HBV model, a ~~single~~ simple proportionality coefficient to estimate the magnitude of the snowmelt is used. This coefficient, multiplied by a constant (usually set to 0.05 in HBV), is also used to compute refreezing rates. ~~However~~Nevertheless, while the degree-day factor is often assumed to be constant over time, seasonal ~~there are good reasons to assume temporal variations due to~~ changes ~~such as~~in snow albedo and solar inclination point out to temporal variations of the degree-day factor as well. While some models use monthly values for this parameter (Quick and Pipes, 1977), ~~A~~a more elegant but still simple way to represent this variability is to consider ~~the~~ a seasonally variable degree-day factor ~~to be seasonally variable~~ following a sine function defined by a yearly average degree-day factor parameter, C_0 *[mm ~~/~~Δt*$^{-1}$ *°C*~~$^{-1}$~~*]*, and an amplitude parameter, $C_{0,a}$ *[mm ~~/~~Δt*$^{-1}$ *°C*~~$^{-1}$~~*]*, defining the amplitude of the seasonal variation (Eq. 11) (Braun and Renner, 1992; Hottelet et al., 1994). By establishing a seasonally-variable degree-day factor instead of a constant value for this parameter, potential snowmelt rates ~~are~~ become smaller during the winter months, ~~while~~ and increa~~s~~ses~~ing~~ during spring and summer (if there is any snow left).

$$C_{0,n} = C_0 + \frac{1}{2} C_{0,a} \sin \frac{2\pi(n-81)}{365}, \tag{11}$$

2.2.5 Snowmelt and Rrefreezing

All liquid water produced by sSnowmelt water does not leave the snowpack directly,; as a certain amount of liquid water may can be stored in the snow. This is important as it thus delayings the outflow of water from the snowpack., and besides thatThe, the liquid water stored in the snowpack can also potentially refreeze if temperatures decrease againbelow the freezing point. In the HBV model, both the storage of liquid water and refreezing processes are considered. However, since the magnitude of refreezing meltwater is generally tiny compared to other fluxes, some models disregard this process entirely as it adds complexity to the model without adding any value to itto reduce model complexity (Magnusson et al., 2014). Here we follow the approach by Magnusson et al. (2014) which, besides disregarding the refreezing process, describes the snowmelt magnitude using an exponential function (Eq. 12). This conceptualisation formulation of snowmelt isf somewhat more realisticdetailed than the one used in HBV butand requires the use of an additional parameter to control for the smoothness of the snowmelt transition, M_M *[°C]*. This way, and cContrary to the formulation used in the standard HBV model, snowmelt occurs even below the freezing point, but at negligible amounts. The impact of increasing temperatures on snowmelt is also higher for this formulation compared to HBVconceptualisation.

$$M = C_0 \cdot M_M \left[\frac{T - T_T}{M_M} + \ln\left(1 + e^{-\frac{T-T_T}{M_M}}\right)\right] \left(\frac{T - T_T}{M_M} + \ln\left(1 + e^{-\frac{T-T_T}{M_M}}\right)\right),$$

(121)

2.3 Study Ddomain and Ddata

We selected two sets of mountainous catchments located atin two different geographical domainscountries within Central Europe to test the proposed modifications to the individual components of the snow routine of the HBV model (Table 2, Figure 1). The first set, composed of Swiss catchments, contains a range of catchments ranging from high--altitude, steep catchments in the central Alps to low-altitudeer catchments in the Pre-Alps and Jura mountains with gentler topography. The second set, composed of Czech catchments, is representative of mountain catchments at lower elevations compared to the Swiss catchments.

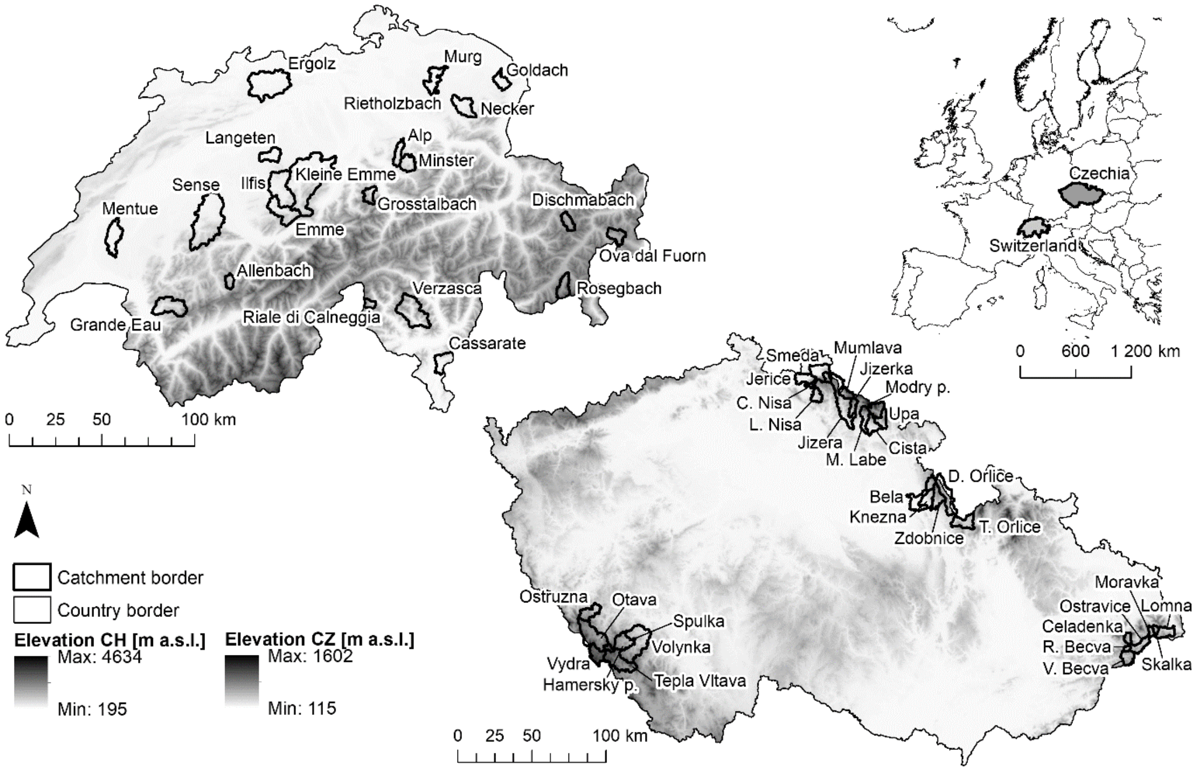


Figure 1 Geographical location of the catchments used in this study. We used a total of 54 catchments; 22 located in Switzerland and 32 in Czechia.

Table 2 Relevant physical characteristics of the catchments included in the study. Each catchment is given an identification code in the following way: country (CH – Switzerland, CZ – Czechia), geographical location (Switzerland: 100 – Jura and Swiss Plateau, 200 – Central Alps, 300 – Southern Alps; Czechia: 100 – Bohemian Forest, 200 – Western Sudetes, 300 – Central Sudetes, 400 – Carpathians), and a sequential number for increasingly snow-dominated catchments within each geographical setting. The official hydrometric station IDs from FOEN and CHMI are also provided.

ID	Catchment	Station	Station ID	Area [km^2]	Mean elevation [m a.s.l.]	Elevation range [m a.s.l.]	Snowmelt contribution to runoff [%]
CH-101	Ergolz	Liestal	2202	261.2	604	305 – 1087	5
CH-102	Mentue	Yvonand	2369	105.3	690	469 – 915	5
CH-103	Murg	Wängi	2126	80.1	657	469 – 930	7
CH-104	Langeten	Huttwil	2343	59.9	770	632 – 1032	8

CH-105	Goldach	Goldach	2308	50.4	825	401 – 1178	14
CH-106	Rietholzbach	Mosnang	2414	3.2	774	697 – 868	9
CH-107	Sense	Thörishaus	2179	351.2	1091	551 – 2096	12
CH-108	Emme	Eggiwil	2409	124.4	1308	770 – 2022	22
CH-109	Ilfis	Langnau	2603	187.4	1060	699 – 1973	14
CH-110	Alp	Einsiedeln	2609	46.7	1173	878 – 1577	19
CH-111	Kleine Emme	Emmen	2634	478.3	1080	440 – 2261	16
CH-112	Necker	Mogelsberg	2374	88.1	970	649 – 1372	16
CH-113	Minster	Euthal	2300	59.1	1362	891 – 1994	26
CH-201	Grande Eau	Aigle	2203	131.6	1624	427 – 3154	26
CH-202	Ova dal Fuorn	Zernez	2304	55.2	2359	1797 – 3032	36
CH-203	Grosstalbach	Isenthal	2276	43.9	1880	781 – 2700	28
CH-204	Allenbach	Adelboden	2232	28.8	1930	1321 – 2587	38
CH-205	Dischmabach	Davos	2327	42.9	2434	1657 – 3024	52
CH-206	Rosegbach	Pontresina	2256	66.6	2772	1771 – 3793	62
CH-301	Riale di Calneggia	Cavergno	2356	23.9	2079	881 – 2827	42
CH-302	Verzasca	Lavertezzo	2605	185.1	1723	546 – 2679	27
CH-303	Cassarate	Pregassona	2321	75.8	1017	286 – 1904	4
CZ-101	Vydra	Modrava	135000	89.8	1140	983 – 1345	34
CZ-102	Otava	Rejstejn	137000	333.6	1017	598 – 1345	29
CZ-103	Hamersky potok	Antygl	136000	20.4	1098	978 – 1213	26
CZ-104	Ostruzna	Kolinec	139000	92.0	755	541 – 1165	17
CZ-105	Spulka	Bohumilice	141700	104.6	804	558 – 1131	19
CZ-106	Volynka	Nemetice	143000	383.4	722	430 – 1302	17
CZ-107	Tepla Vltava	Lenora	106000	176.0	1010	765 – 1314	20
CZ-201	Jerice	Chrastava	319000	76.0	493	295 – 862	14
CZ-202	Cerna Nisa	Straz nad Nisou	317000	18.3	672	368 – 850	13
CZ-203	Luzicka Nisa	Prosec	314000	53.8	611	419 – 835	22
CZ-204	Smeda	Bily potok	322000	26.5	817	412 – 1090	26
CZ-205	Smeda	Frydlant	323000	132.7	588	297 – 1113	18
CZ-206	Jizera	Dolni Sytová	086000	321.8	771	399 – 1404	26
CZ-207	Mumlava	Janov-Harrachov	083000	51.3	970	625 – 1404	34
CZ-208	Jizerka	Dolni Stepanice	086000	44.2	842	490 – 1379	29
CZ-209	Malé Labe	Prosecne	003000	72.8	731	376 – 1378	25
CZ-210	Cista	Hostinne	004000	77.4	594	358 – 1322	19
CZ-211	Modry potok	Modry dul	008000	2.6	1297	1076 – 1489	38
CZ-212	Upa	Horni Marsov	013000	82.0	1030	581 – 1495	28

CZ-213	Upa	Horni Stare Mesto	014000	144.8	902	452 – 1495	25
CZ-301	Bela	Castolovice	031000	214.1	491	269 – 1104	25
CZ-302	Knezna	Rychnov nad Kneznou	030000	75.4	502	305 – 861	25
CZ-303	Zdobnice	Slatina nad Zdobnici	027000	84.1	721	395 – 1092	24
CZ-304	Divoka Orlice	Klasterec nad Orlici	024000	153.6	728	505 – 1078	22
CZ-305	Ticha Orlice	Sobkovice	032000	98.5	622	459 – 965	22
CZ-401	Vsetinska Becva	Velke Karlovice	370000	68.3	749	524 – 1042	22
CZ-402	Roznovska Becva	Horni Becva	383000	14.1	745	568 – 966	24
CZ-403	Celadenka	Celadna	279000	31.0	803	536 – 1187	30
CZ-404	Ostravice	Stare Hamry	275300	73.3	707	542 – 922	32
CZ-405	Moravka	Uspolka	281000	22.2	763	560 – 1104	30
CZ-406	Skalka	Uspolka	282000	18.9	785	571 – 1029	24
CZ-407	Lomna	Jablunkov	298000	69.9	667	390 – 1011	25

2.3.1 Switzerland

We selected 22 catchments in Switzerland covering a wide range of elevations and areas in the three main hydro-geographical domains of the country, i.e. the Jura and Swiss Plateau, the Central Alps, and the Southern Alps (Weingartner and Aschwanden, 1989). No~~The choice of catchments was constrained by our intention to avoid~~ catchments with significant karst or glacierised areas, as well as catchments with substantial human influence on runoff were selected for this study. This decision allowed us to observe the signal of snow processes, without including noise or added complexity from other processes, but limited the number of catchments in high altitudes, which are the ones with largest snowmelt contribution to runoff~~, and therefore those that would potentially benefit the most from an increased realism of the snow routine of the model~~. The resulting set of catchments had~~s~~ mean elevations between 600 and 2800 m a.s.l. with elevation gradients of up to 2000 m and catchment areas between 3 and 500 km^2 (Figure 2). There was~~is~~ a ~~great~~ considerable variability in the yearly snowmelt contribution to runoff, ranging from 5 % to 60 % as the catchments ranged ~~between~~ from pluvial to glacio-nival regimes.

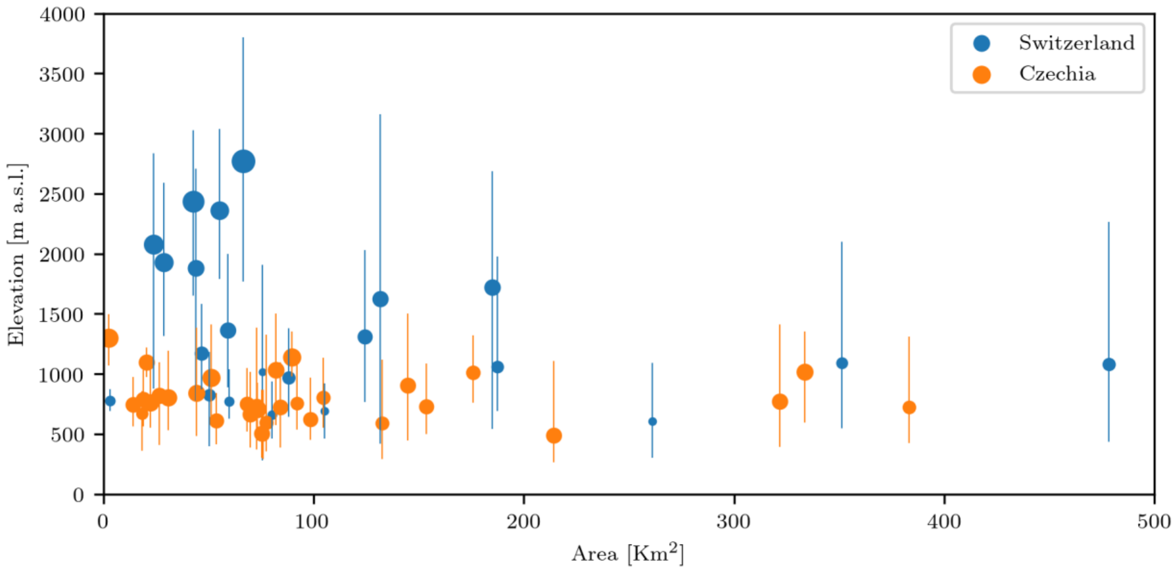
We obtained the necessary meteorological data for running the HBV model from the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss). More specifically, we used pre-processed gridded data products to obtain catchment-average precipitation (Frei et al., 2006; Frei and Schär, 1998), and temperature (Frei, 2014). These gridded data products are available from 1961, have a daily temporal resolution, and a spatial resolution of 1.25 degree minutes covering the entire country.

We used both stream runoff and snow water equivalent data for model calibration and validation. We obtained daily stream runoff data from the Swiss Federal Office for the Environment (FOEN, 2017). Regarding snow water equivalent, we used 18 years of gridded daily snow water equivalent data at 1 km^2 resolution derived from a temperature-~~-~~index snow model with integrated three-dimensional sequential assimilation of observed snow data from 338 stations of the snow monitoring networks of MeteoSwiss and the Swiss Institute for Snow and Avalanche Research (SLF) (Griessinger et al., 2016; Magnusson et al., 2014). Even if using a temperature-index model for both the HBV model and the estimation of the snow water equivalent validation data may introduce some bias to the results, the data assimilation and error correction methods used in the estimation of snow water equivalent make this methodology especially robust (Magnusson et al., 2014). Finally, we obtained the catchment areas and topography from a digital elevation model with a resolution of 25 m from the Swiss Federal Office of Topography (swisstopo, https://swisstopo.admin.ch/).

2.3.2 Czechia

The second set of catchments was composed of Czech catchments and includes 32 mountain catchments with catchment areas ranging from 3 to 383 ~~–~~km^2 (Figure 2). As for Switzerland, w~~W~~e selected near-natural catchments with no major human influences such as big dams or water transfers. The ~~resulting~~ selected catchments ~~are~~ were located at ~~relatively~~ lower elevations ~~and present lower elevation ranges compared to~~than most of the selected Swiss catchments. Additionally, they ~~are~~ were located in the transient zone between oceanic and continental climate, with lower mean annual precipitation ~~compared to~~than the Swiss catchments. The mean annual snow water equivalent ~~maximum~~ peak for the period 1980 – 2014 ranged~~s~~ from 35 mm to 742 mm depending on catchment elevation, resulting in 13 % to 39 % of the annual runoff coming from spring snowmelt.

We obtained daily precipitation, daily mean air temperature, and daily mean runoff time series from the Czech Hydrometeorological Institute (CHMI). Additionally, we obtained weekly snow water equivalent data from CHMI (measured each Monday at 7 CET). Since no gridded ~~data of~~ precipitation, air temperature, or snow water equivalent data ~~are~~are available for Czechia, station data were used for HBV model parametrization. We used stations located within the individual catchments when available. ~~When~~ If no such station was available, we selected the nearest station representing similar ~~catchment~~ conditions to the target catchment (e.g., stations situated at a similar elevation). Finally, we used a digital elevation model with a vertical resolution of 5 m from the Czech Office for Surveying, Mapping and Cadastre to obtain catchment areas and ~~topography~~elevation distributions.



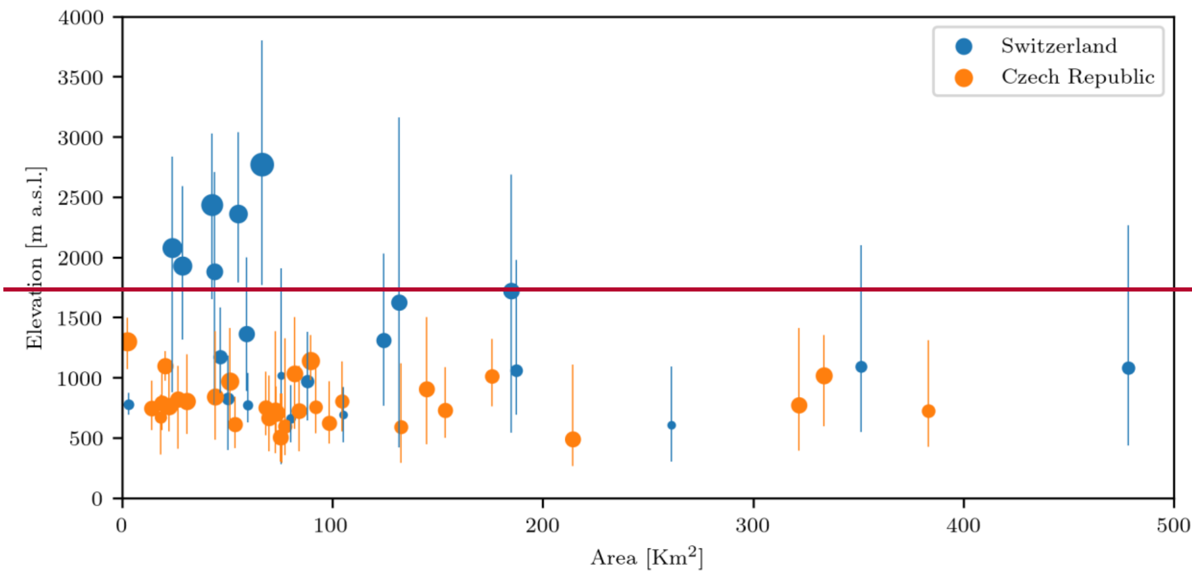


Figure 2 Distribution of the catchments used in this study ~~regarding~~ in terms of~~their~~ area (x-axis), mean elevation ~~mean value~~ and range (y-axis), and relative snowmelt contribution to runoff (marker size). The catchments are coloured according to their respective geographical domain: blue (Switzerland), and orange (Czechia).

2.4 Experimental s~~S~~etup

Even if sub-daily data were available for most variables for the Swiss catchments, we considered that daily data was ~~suitable~~ beneficial for this study, as using sub-daily temporal resolutions would have required to tak~~eing~~ into account the diurnal variability of some of the variables, thus requiring a higher ~~model complexity~~ comprehensiveness over the included hydrological processes in the model (Wever et al., 2014). For instance, radiation and temperature fluctuations along the day would require similarly variable degree-day factor values (Hock, 2005). Other factors such as ~~travel~~ the transport time~~s~~ of meltwater from the snowpack to the streams ~~between the sources of snowmelt and the streams~~ would also become ~~significant issues~~relevant at sub-daily time scales (Magnusson et al., 2015). ~~In order t~~To keep the model simple but at the same time being able to represent the elevation-dependent snow processes, we used a single vegetation zone per catchment but divided the catchment area into 100 m elevation zones (Uhlenbrook et al., 1999).

When evaluating the performance of ~~hydrological~~ rainfall-runoff models to simulate snow dynamics, this evaluation is sometimes done solely ~~looking at~~against ~~the simulated~~ runoff ~~by the model~~observations, as this variable is the main output of ~~hydrological~~ such models (Riboust et al., 2019; Watson and Putz, 2014). Nevertheless, this analysis alone is incomplete as the performance of the model to reproduce runoff is the result of the interaction between the different routines and components of the model, also those that are not directly related to snow processes. A direct evaluation of the relevant model routine (i.e. the snow routine in this case) should be performed as well. Focusing on the snow routine, snow cover fraction and snow water equivalent are widely-adopted evaluation metrics (Avanzi et al., 2016; Helbig et al., 2015). The fact that snow water equivalent is a more direct measure of the amount of water that will eventually be converted to runoff, in addition to the difficulties in accurately determining the snow cover fraction for our study area and period, led us to choose snow water equivalent for evaluating the snow routine structure of the model. ~~For this reason~~In short,~~,~~ ~~here~~ we evaluated the ~~existing~~ different model structures ~~as well as the proposed modifications to it~~ based on their ability to represent (i) the snow water equivalent of the snowpack, and (ii) stream runoff at the catchment outlet.

To evaluate the performance of the different model structures to reproduce the snow water equivalent of the snowpack, we used a modified version of the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) where the model performance, R_W, is given by the fraction of the sum of quadratic differences between snow water equivalent observations, W_O, and simulations, W_S, and between observations and the mean observed value, $\overline{W_o}$ (Eq. 13~~2~~).

$$R_W = 1 - \frac{\sum(W_o - W_S)^2}{\sum(W_o - \overline{W_o})^2}, \qquad (13~~2~~)$$

Due to the substantial differences data availability regarding in snow water equivalent (SWE) values ~~data availability~~ between the two datasets (gridded data in Switzerland vs point data in Czechia), we had to adapt the model calibration and evaluation procedure to each case. We evaluated the model against the mean snow water equivalent value for each elevation zone for the Swiss catchments, and against the measured values at a given elevation for the Czech ones.

Regarding the evaluation of the model against stream runoff ~~estimation~~, we deemed that the standard Nash-Sutcliffe efficiency measure was not suitable for our case study as it is skewed towards high flows (Schaefli and Gupta, 2007). Snow processes are dominant both in periods of high flows (e.g. spring flood) and low flows (e.g. winter conditions), which are equally important for our purposes ~~to estimate correctly~~. For this reason, we decided to evaluate the estimation of stream runoff by using the natural logarithm of runoff instead (Eq. 14~~3~~).

$$R_{\ln Q} = 1 - \frac{\sum(\ln Q_o - \ln Q_s)^2}{\sum(\ln Q_o - \overline{\ln Q_o})^2}, \qquad (14~~3~~)$$

Some studies focusing on snow hydrology establish specific calibration periods for each catchment based on, for instance, the snowmelt season (Griessinger et al., 2016). In this study, however, we decided to constrain the calibration and evaluation periods in a consistent and automated manner for all catchments. For this reason, we defined the model calibration and evaluation periods as comprising days with significant snow cover on the catchment (>25% of the catchment covered by snow). We ~~also considered~~included a full week after the occurrence of snowmelt ~~in order~~ to account for runoff delay. We obtained the value of 25% through empirical tests on the number of days with specific snow coverage values and their corresponding snow water equivalent values for each catchment. We found that below this value the total snow water equivalent in the studied catchments usually becomes negligible.

We calibrated the model for all the catchments in the study using a split-sample approach. We selected this approach because it allowed us to assess both the best possible model ~~efficiency~~ performance with respect to each objective function for each model structure variant~~alternative~~ (i.e. calibration period), and a realistic model application scenario (i.e. validation period), helping us to distinguish between real model improvement and overfitting. In our case, the simulation period was limited by the input data with the shortest temporal availability, which in this case was the snow water equivalent data for the Swiss catchments. In total 20 years were available, which we divided into two equally long 9-year periods plus 2 years for model warm-up. We calibrated the model for both periods and cross-validated the simulations ~~for~~on the remaining periods. For the Swiss catchments, we ~~considered~~used the period between 1st of September 1998 and 31st August 2016, while for the Czech catchments we ~~considered~~used the period between 1st November 1996 and 31st October 2014. The different start dates for simulation periods in the Swiss and Czech catchments correspond to the different timing for the onset of snow conditions in the different areas. Additionally, the different years included in each study domain correspond to data limitations in each area. Since ~~both~~ the two areas were quite distant~~are geographically separated~~, we considered that it was more important to have the same period length for running the simulations in both domains rather than using the exact same years, as the meteorological conditions are different in the two study domains anyway.

~~Overall, w~~We calibrated the model for all possible combinations of the single modifications to individual components of the snow routine of the HBV model described in Section ~~0~~2.2 (n = 64), catchments (n = 54), simulation periods (n = 2), and objective functions (n = 2) using a genetic algorithm (Seibert, 2000). Every calibration effort consisted of 3500 model runs with constrained parameter ranges based on previous studies (Seibert, 1999; Vis et al., 2015). We performed ~~10~~ten independent calibrations for each setup ~~in order~~ to be able to capture the uncertainty of the model. In total we performed ~~around~~approximately 500 million model simulations. To to assess the impact of potential equifinality and parameter uncertainty issues, we performed a Monte Carlo sensitivity analysis on the all calibration parameters for each for the model structure variants.

3 Results

The large number of catchments and model variations considered in this study madke it difficult challenging to grasp the detailed resultsany details when looking at the entire dataset. ThereforeFor this reason, we first presentdescribe the results for aone single catchment to explore the implications of individual model modifications and illustrate the general trends observed across the study domain. For this purpose we selected the Allenbach catchment at Adelboden (CH-204), one of the high altitude, snow-dominated catchments in the set, as sample catchment. Thereafter, and then we progressively add more elements into the analysis of the results. Additionally, even if we calibrated (and validated) the model for both periods defined in the split-sample test, here we only present the results for the calibration effort in period 1 and corresponding model validation in period 2, as they are representative for the entire analysis. A comprehensive list including calibration and validation model performance values for both objective functions and all catchments included in this study can be found in Appendix A.As an example catchment, we selected one of the high altitude, snow-dominated catchments in the set, the Allenbach catchment at Adelboden (CH-204), as it allows us to describe some of the general trends observed across the study domain.

The calibration performance of the standard HBV model for this the Allenbach catchment is was satisfactoryvery high for both objective functions (model performance values of ~0.90) but, still, some modifications led to increased model performances. Looking atAmongst all the different single modifications tochanges in single components of the snow routine structure of the HBV model that we evaluated in this study, using a seasonally varying degree-day factor ($C_{0,s}$) has had the most substantial impact on the model performance to represent snow water equivalent andfollowed by, to a lesser extent, stream runoff (Figure 3). Apart from thisese modifications, only the use of an exponential function to define the precipitation partition between rain and snow (ΔP_e) produces produced significant changes in the model performance against both objective functions. In this case, however, this modification impacts impacted the model performance in different opposite ways, depending on the objective function, even leading to decreased model performance when for the calibrationg against stream runoff. Finally, if we look at the model, we observe that mModel uncertainty, as given by the performance ranges obtained when aggregating the different calibration efforts, is was small when compared to the performance differences between the different model structures.

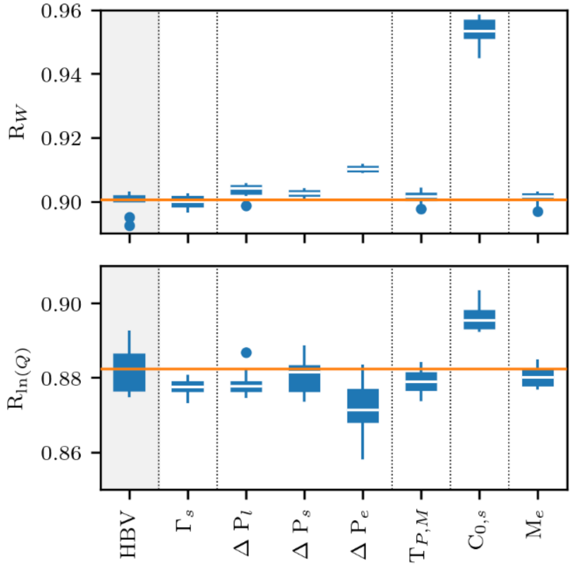
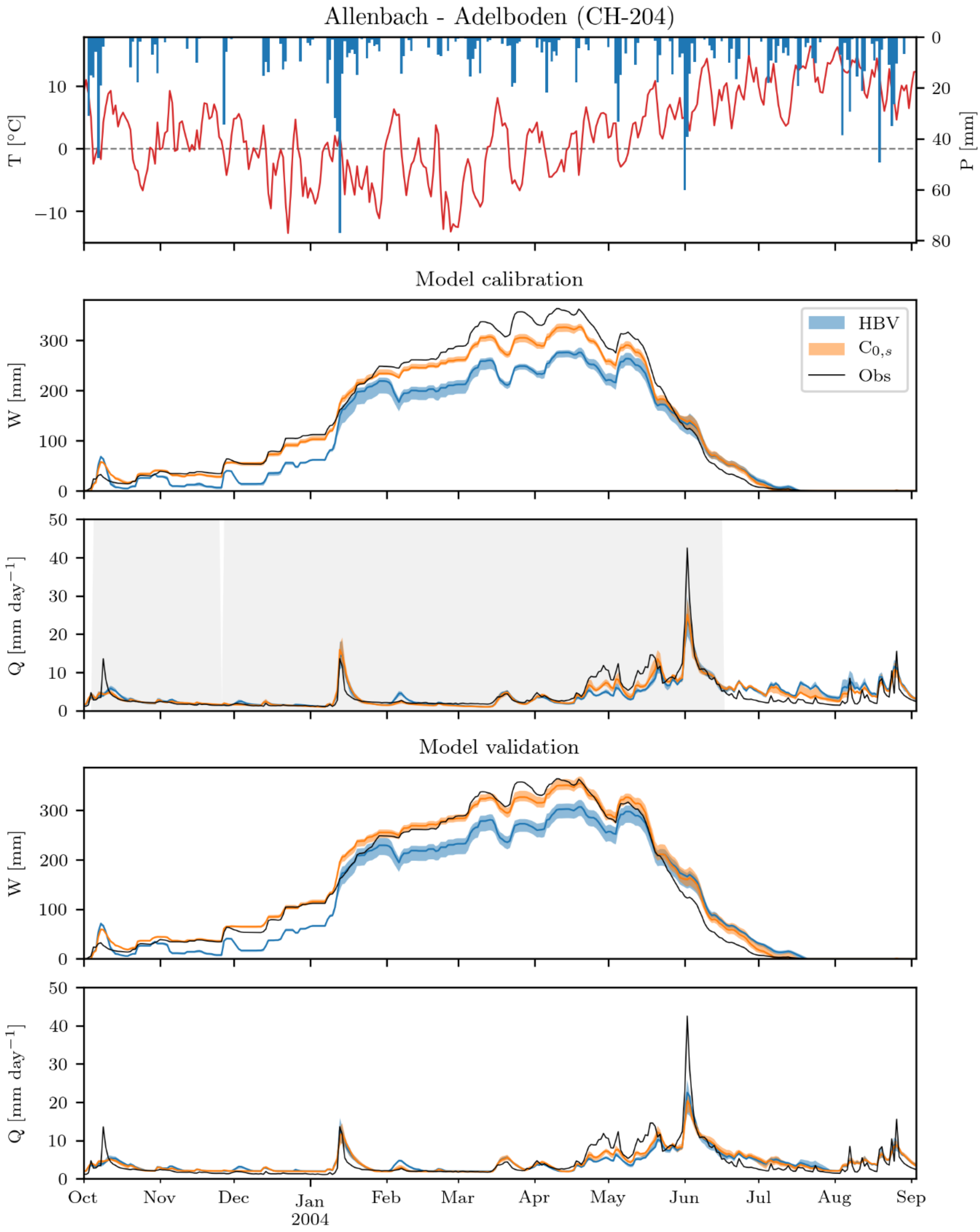


Figure 3 Model calibration performance for the 10 calibration efforts against the two objective functions (top: snow water equivalent; bottom: logarithmic stream runoff) for each of the modifications to individual components of the snow routine of the HBV model for the Allenbach catchment at Adelboden (CH-204). The modifications include a seasonally-variable temperature lapse rate (Γ_s), a linear, sinusoidal, and exponential function for the precipitation phase partition (ΔP_l, ΔP_s, and ΔP_e respectively), different thresholds for precipitation phase and snowmelt ($T_{P,M}$), a seasonally variable degree-day factor ($C_{0,s}$), and an exponential snowmelt with no refreezing (M_e). The median efficiency performance of HBV is represented with an orange horizontal line.

If we lLooking at an example sample year within the calibration period, we can get a grasp onf how comparable the simulated values of snow water equivalent and stream runoff, (including the model uncertainty,) comparedare to with the observed values (Figure 4Figure 3). While capturing the general evolution of the snowpack, the HBV model tends tended to underestimate the snow water equivalent amounts, except for the spring snowmelt period. The model alternative using a seasonal degree-day factor ($C_{0,s}$), which had , as we have already seen, isproven to be the best possible model structure modification for model calibration against snow water equivalent for this catchment, exhibits exhibited the same overall behaviour but is being more accurate and precise than the HBV model structure. Regarding the calibration against stream runoff, both model alternatives performed well for low flow periods, but they missed or underestimated some of the peaks. Model uncertainty is was comparable for both model alternatives and is was not significant when compared to the simulated values. Model results from the same sample year for all the catchments included in this study can be found in the Supplement.


Allenbach - Adelboden (CH-204)
T [°C]
P [mm]
Model calibration
W [mm]
HBV
$C_{0,s}$
Obs
Q [mm day^{-1}]
Model validation
W [mm]
Q [mm day^{-1}]
Oct
Nov
Dec
Jan
2004
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep

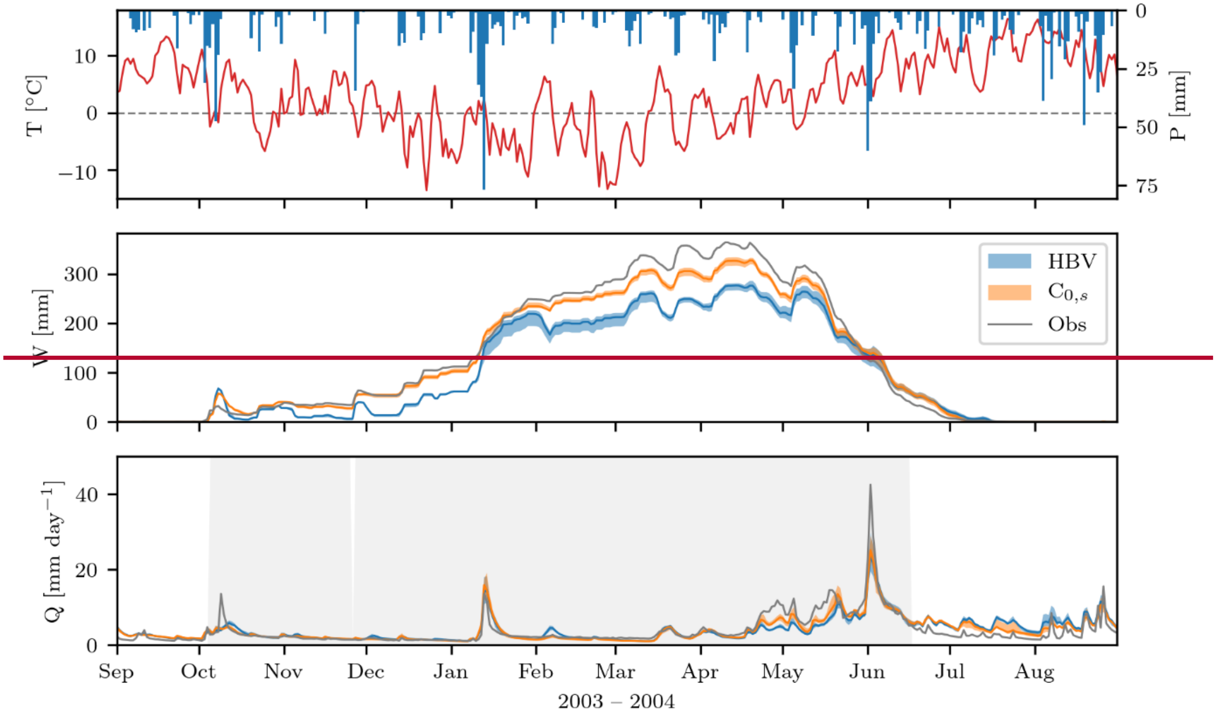


Figure 4 Example time series (September 2003 – August 2004) from the Allenbach catchment at Adelboden (CH-204)~~ within the calibration period~~. Top: daily mean air temperature and total precipitation. Middle: ~~catchment-average observed (grey line) and simulated snow water equivalent (HBV in blue and the model structure modification including a seasonally-varying degree-day factor, $C_{0,s}$ in orange)~~model calibration results. Bottom: ~~observed (grey line) and simulated stream runoff (HBV in blue and the model structure modification including a seasonal degree-day factor in orange)~~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.

~~Zooming out to~~Looking at the entire set of catchments, ~~we can observe that, for the calibration period, ~~the impact of the different model structure modifications on ~~the ~~model calibration performance ~~is ~~was generally more pronounced for R_W than for $R_{\ln(Q)}$ across all ~~of the considered ~~catchments (Figure 5). For most catchments, ~~most ~~the largest model performance improvements when calibrating against snow water equivalent were achieved by using a seasonally variable degree-day factor ($C_{0,s}$). Using different thresholds for precipitation phase partition and snowmelt ($T_{P,M}$) and using an exponential function for precipitation phase partition (ΔP_e) also conveyed a significant improvement~~s~~ ~~in ~~for some of the catchments. Nevertheless, ~~this last~~the latter modification ~~performs ~~performed almost equal to the HBV model when calibrating against stream runoff, and even slightly worse for some catchments~~ even slightly worse~~. Using an exponential function to define the precipitation partition between rain and snow consistently ~~penalises ~~penalised the model performance when calibrating the HBV model against stream runoff, whereas using an exponential function for snowmelt (M_e) ~~is ~~was the best alternative when ~~looking at~~calibrating the model against this objective function. Overall, most modifications conveyed slight model performance improvements ~~in model performance with respect to~~concerning snow water equivalent simulations for most of the catchments in the dataset. Nevertheless, the modifications on the precipitation phase partition tended to penalise most Czech catchments when calibrating against snow water equivalent. We did not observe any significant connection between model performance and catchment characteristics such as mean catchment elevation, catchment area, or yearly snowmelt contribution to runoff.

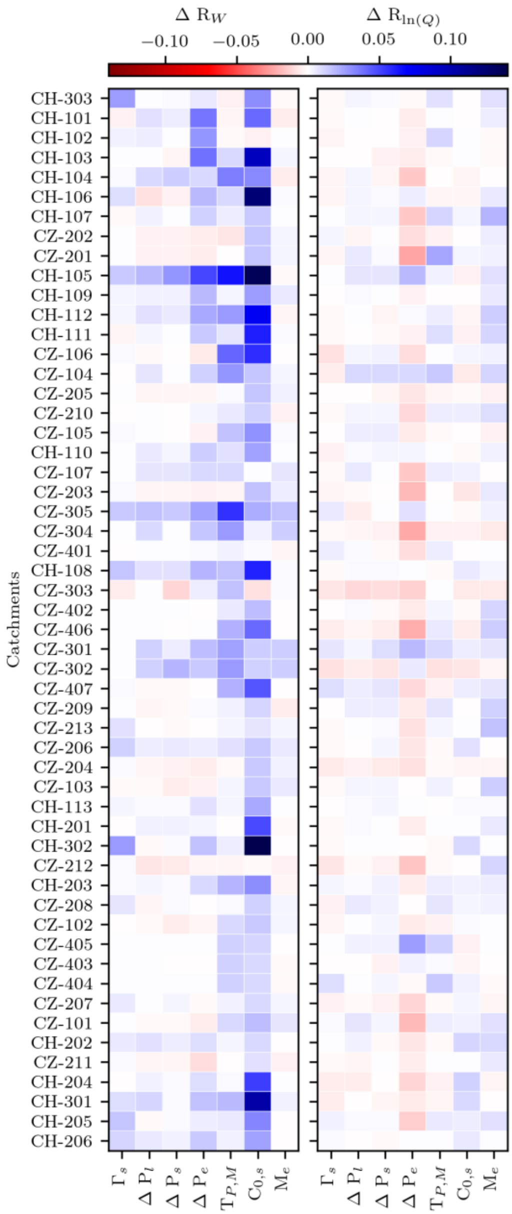


Figure 5 Median relative model calibration performance for alternative HBV model structures, including modifications to single components of the snow routine with respect to HBV. The modifications include a seasonally-variable temperature lapse rate (*Γ_s*), a

linear, sinusoidal, and exponential function for the precipitation phase partition (ΔP_l, ΔP_s, and ΔP_e respectively), different thresholds for precipitation phase and snowmelt ($T_{P,M}$), a seasonally variable degree-day factor ($C_{0,s}$), and an exponential snowmelt with no refreezing (M_e). Left: model calibration against snow water equivalent; right: model calibration against logarithmic stream runoff. The catchments are ordered by mean yearly snowmelt contribution to runoff in downwards increasing order.

~~So far, we have seen that w~~While some modifications ~~have~~ had a clear and consistent impact on model calibration performance in all catchments, most of them ~~have~~ presented a less pronounced~~, catchment-dependent~~ either positive or negative impact, depending on the catchment, making it difficult to evaluate ~~. It is therefore difficult to decide~~ which ~~modifications are better~~ model structures were more suitable than others (including the default HBV structure) for most of the catchments. Additionally, ~~Furthermore, until this point, we have only looked at the calibration efficiencies. T~~to better understand the usefulness of the different modifications in real applications, we need to take into account which of ~~the modifications~~model structures performed best ~~with respect to~~for the validation period as well (Figure 6). As already observed in Figure 5, using a seasonal degree-day factor ($C_{0,s}$) ~~is~~ was the best modification for calibrating the model against snow water equivalent for the vast majority of the catchments. Nevertheless, this modification ~~ranks~~ ranked relatively low ~~for~~ when validating the model~~ validation~~ against the same objective function. Looking at stream runoff, using an exponential function for snowmelt simulation while disregarding the refreezing process (M_e) ~~is~~ was the best-ranking modification for both model calibration and validation while the HBV model~~ structure~~ ~~ranks~~ ranked higher than several of the considered modifications. Using an exponential function to define the precipitation partition between rain and snow (ΔP_e) ~~is~~ was the worst alternative for calibrating ~~it~~the model against stream runoff. ~~For model calibration, t~~There ~~is a~~ diagonal pattern from the top left to the bottom right observed for model calibration ~~(Figure 6),~~ indicat~~es~~ing that ~~different~~ modifications tended to have the same rank ~~in~~for most catchments (not~~ic~~e that the ranking of modifications is different when looking at snow water equivalent with respect to stream runoff). Such a pattern ~~is~~ was not present for model validation, suggesting that, in that case, there ~~is~~ were no ~~clear answer to which modifications~~ model structures significantly more suitable than others~~convey most value to the model~~.

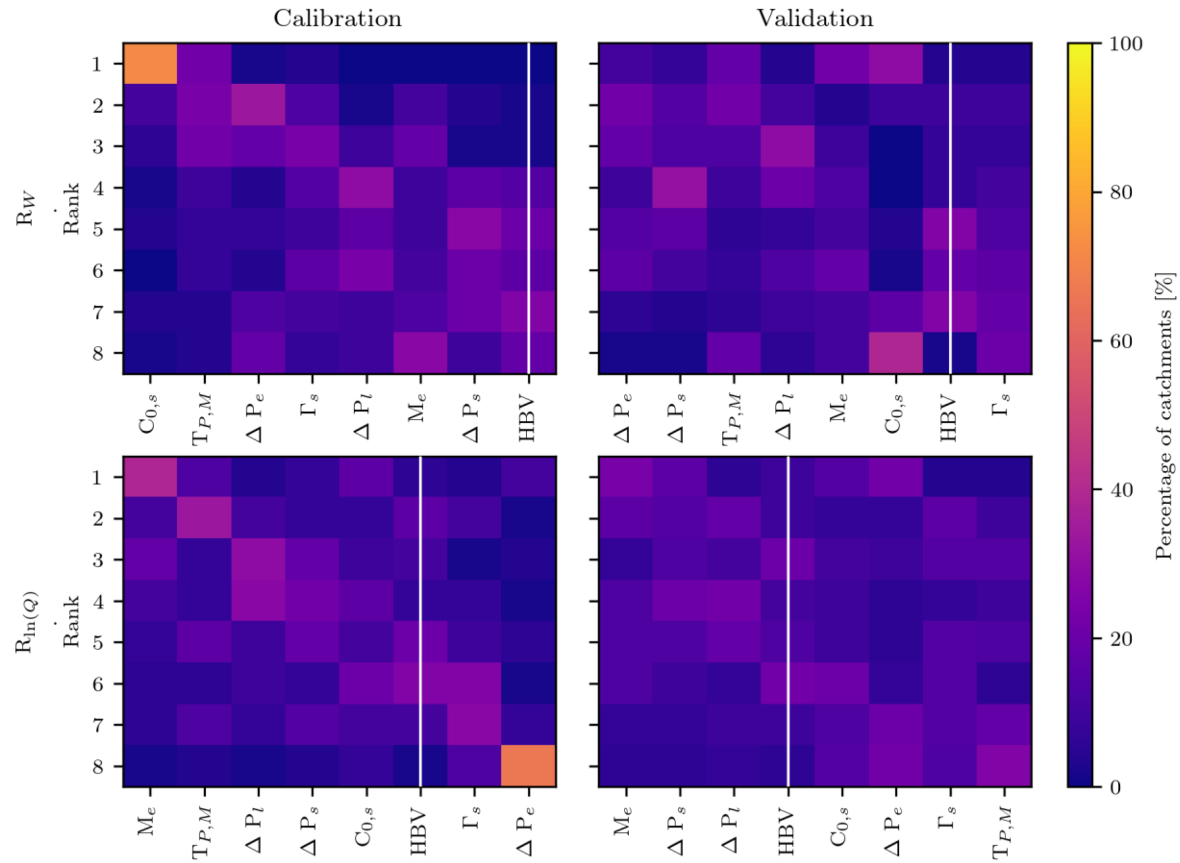


Figure 6 Rank matrices for each of the model simulation scenarios. Top: model calibration (left) and validation (right) against snow water equivalent; bottom: model calibration (left) and validation (right) against logarithmic stream runoff. Each rank matrix shows the rank distribution of each modification to single components of the snow structure of the HBV model for all the catchments included in this study so that each column adds to 100%. The modifications are ordered from highest to lowest average ranking (left to right) and include a seasonally-variable temperature lapse rate (Γ_s), a linear, sinusoidal, and exponential function for the precipitation phase partition (ΔP_l, ΔP_s, and ΔP_e respectively), different thresholds for precipitation phase and snowmelt ($T_{P,M}$), a seasonally variable degree-day factor ($C_{0,s}$), and an exponential snowmelt with no refreezing (M_e). The HBV model structure is highlighted with a white vertical line.

~~As we have seen, e~~Even if some ~~model modifications~~alternative model structures clearly improved the calibration performance of the model~~ for model calibration~~, most ~~modifications~~ structures ~~have~~ had a limited~~at most minimal~~ impact on model performance. ~~A reason for t~~This ~~might~~is in part because, to this point, we only tested model structures containing a single modification with respect to the HBV model~~ be that the actual differences compared to the standard HBV model formulation are minimal~~. ~~For this reason, w~~We next explored whether the same trends persisted when ~~increasing~~ including further elements to the model by using ~~the model complexity through incorporating~~ an increasing amount of model structure modifications simultaneously. In, total we tested 64 different model structures ~~The number of model alternatives for each number of modifications to the model structure is presented in~~ (Table 3). Using the maximum possible number of simultaneous modifications (5) to the model structure would result in the use of up to nine additional parameters. This would lead to a clear overparameterisation of the model (the default snow routine structure of HBV contains five parameters), but we included this alternative to provide a complete analysis of the all available alternatives.

Table 3 Number of model structure alternatives containing a given number of snow routine modifications.

Number of modifications	Number of alternatives
0	**1**
1	7
2	18
3	22
4	13
5	3
	n = 64~~3~~

~~For instance, in the case of introducing 5 modifications to the snow routine of the model, the only available alternative representation for the lapse rate (i.e. F_s), the threshold temperature (i.e. $T_{P,M}$), the degree-day factor (i.e. $C_{0,s}$), and snowmelt and refreezing (i.e. M_e) are used, in combination with one of the three alternative representations for the precipitation phase partition (i.e. ΔP_l, ΔP_s or ΔP_e).~~

Figure 7 shows the median ~~relative efficiency~~ model performance for~~of~~ each of the 64 possible model structure alternatives for all catchments relative to the standard HBV model performance, sorted by the number of components being modified. W~~e can see that, w~~hen calibrating the model against snow water equivalent, model ~~efficiency~~ performance clearly ~~increases~~ increased for all of the model structure alternatives. The impact ~~is~~ was more modest for model validation with a significant percentage of alternative structures performing worse than HBV. Regarding model calibration against stream runoff, the effect of an increasing number of components being modified ~~is~~ was ~~minimal~~ limited but mostly positive. The range of ~~efficiencies~~ model performance values ~~is~~ was also significantly smaller than when looking at snow water equivalent. This relates~~, which is due~~ to the fact that, for most catchments, the snow routine has a limited weight over the entire HBV model. For model validation we observed a similar trend, but with broader ~~efficiency~~ model performance ranges. Also, the fact that performance ~~ranges~~ variability varied~~y~~ significantly ~~across~~with the number of components being modified ~~is~~ was in part due to the differences in the number of model structure alternatives for each of them, being larger for those number of modifications which included the largest amount of model structure alternatives.

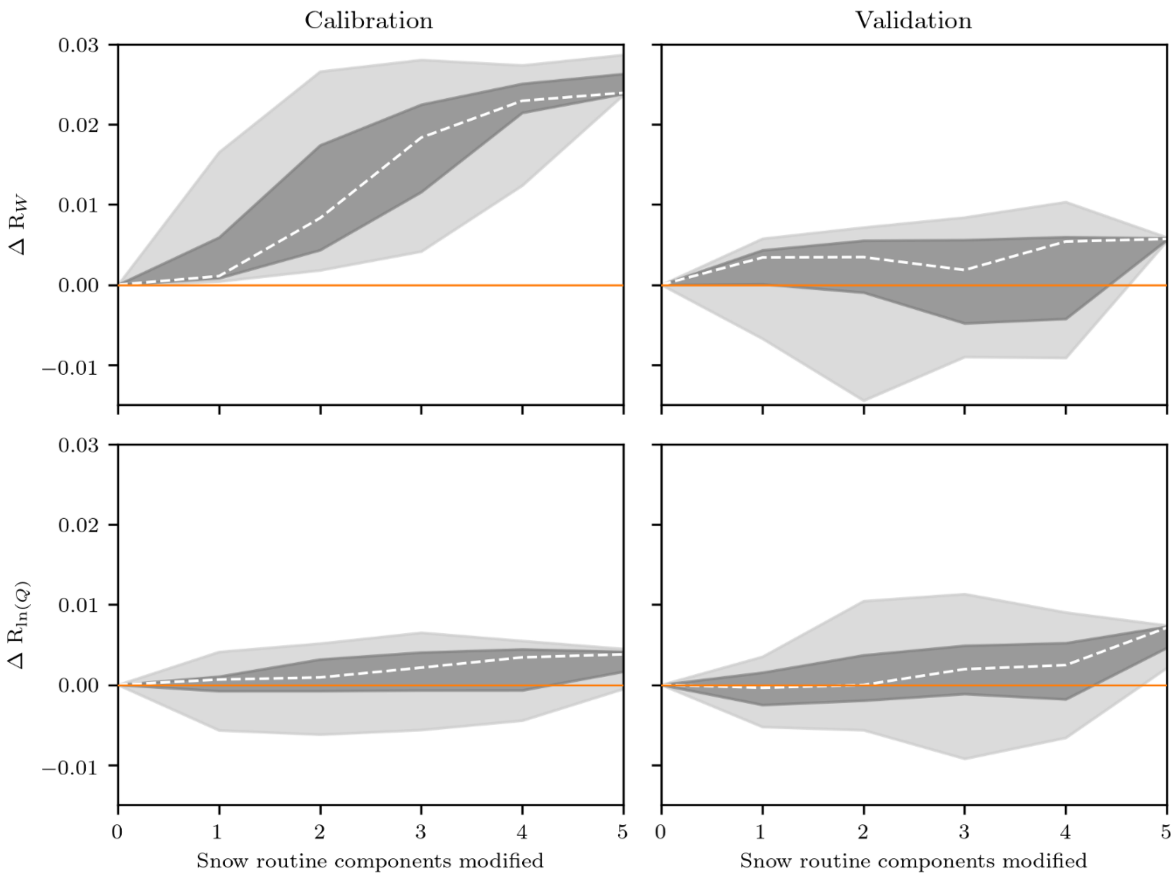


Figure 7 Median relative model ~~efficiency~~ performance with respect to HBV across all catchments for the 64 considered snow routine structures sorted by an increasing number of modifications. Model simulations against snow water equivalent are presented in the top row, while those against logarithmic stream runoff are shown in the bottom row. Model calibration is presented in the left column and model validation in the right one. The dashed line represents the median value across all catchments while the grey fields represent the minimum to maximum (light grey) and 25th to 75th percentiles (dark grey). The relative ~~efficiency~~ model performance of HBV is highlighted with a solid orange line.

In general, ~~there~~ we observed~~is~~ an increase in~~of~~ model ~~efficiency~~ performance for all cases. However, ~~with the exception of~~except for model calibration against snow water equivalent, there ~~is~~ was no clear indication that a model with a more detailed formulation of the snow processes ~~an increasing model structure complexity~~ would lead~~s~~ to significantly improved model performance. Indeed, the range of performances among the different model structures was~~is~~ larger than the median net increase. This might be an indication that choosing the right modifications (or combinations of modifications) is more relevant than significantly increasing model ~~complexity~~detail. This way, we attempted to determine ~~if~~ whether ~~there are~~ some specific modifications ~~that~~ conveyed a model performance gain across all model structures in which they were included for ~~the four simulation scenarios~~both model calibration and validation against the two objective functions. ~~W~~To this purpose, we ranked all model structures for model calibration and validation against ~~the two~~ both objective functions and ~~looked at~~visualised the cumulative distribution of each of the individual model modifications (Figure 8). Some of the patterns observed here ~~remain~~

~~the same as~~resemble those that we already observed for single modifications only (Figure 6). For instance, ~~when looking at model calibration against snow water equivalent,~~ all top-ranking model structures ~~use~~ included a seasonally variable degree-day factor ($C_{0,s}$) for model calibration against snow water equivalent. Similarly, ~~when looking at model calibration against stream runoff,~~ all bottom-ranking model structures used an exponential function for precipitation phase partition (ΔP_e) for model calibration against stream runoff. Besides these familiar patterns, ~~O~~other patterns emerged, ~~that~~ which could not be clearly observed when only looking at single modifications ~~emerge here as well~~. Indeed, even if a seasonal degree-day factor ~~performs~~ performed above average ~~for~~ in most cases, this particular modification ~~is~~ was included in all of the bottom-ranking model structures for model validation against snow water equivalent. ~~Also~~Additionally, model structures including an exponential function for snowmelt (M_e) performed above average for all cases, and were even included in almost all the top-ranking model structures~~ones~~.

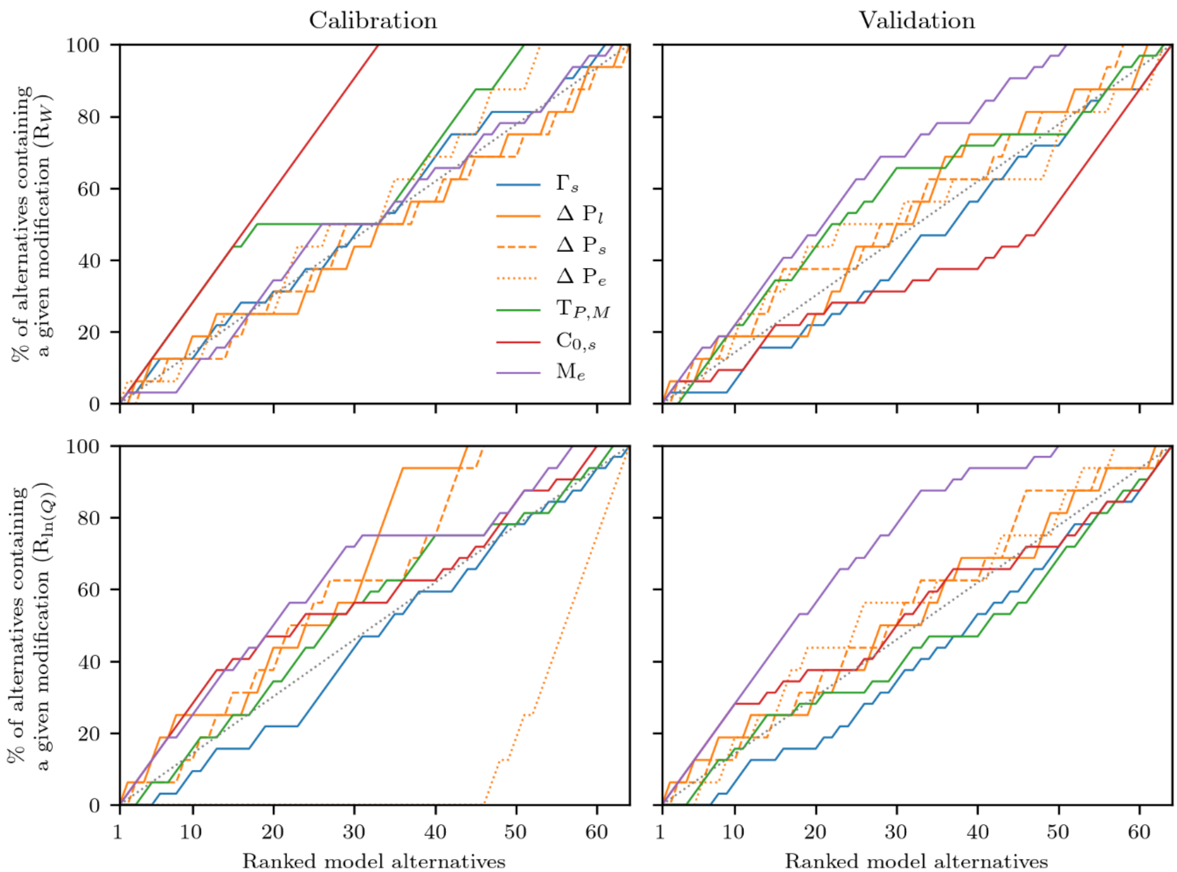


Figure 8 ~~Cummulative~~Cumulative plots for each of the 7 individual modifications to the snow routine of the HBV model as a function of the ranked 64 model structures arising from all the possible combinations of modifications. The modifications are a seasonally-variable temperature lapse rate (Γ_s), a linear, sinusoidal, and exponential function for the precipitation phase partition (ΔP_l, ΔP_s, and ΔP_e respectively), different thresholds for precipitation phase and snowmelt ($T_{P,M}$), a seasonally variable degree-day factor ($C_{0,s}$), and an exponential snowmelt with no refreezing (M_e). Model simulations against snow water equivalent are presented in the top row, while those against logarithmic stream runoff are presented in the bottom row. Model calibration is presented in the ~~l~~left

column and model validation in the right one. Model modifications plotting above the 1:1 line (gr~~a~~ey dotted line) tend to be included in high-ranking model structures, while those plotting below the 1:1 line tend to be included in low ranking structures.

~~Indeed, looking at~~Based on the ranked alternative model structures and the modifications contained in each of them, ~~some~~ specific model modifications contributed the most to model performance increases~~have proven to be dominant~~, ~~even when combined with a number~~ regardless of other model structure modifications. Nevertheless, these dominant modifications impacted the model structure performance in different ways, depending on the modelling scenario (i.e. calibration/validation, objective function). Ideally, ~~the~~ any model structure modifications ~~we are considering to the HBV model~~ should convey an improved representation of snow water equivalent but also have a positive impact on the simulation of stream runoff (which is the main output of the model), both for model calibration and validation.

To achieve an improved representation of both snow water equivalent and stream runoff we ~~should only~~ only took into account~~consider~~ those model ~~alternatives~~structures that ~~have~~led to a positive impact for each of the four modelling scenarios (i.e. calibration and validation efforts against both objective functions). This way, we selected and ranked all model alternatives that ~~have~~ had a positive median relative ~~efficiency~~ model performance value~~s~~ with respect to the HBV model and examined which modifications led to the largest model performance improvement ~~are dominant~~ (Figure 9). All of the selected ~~alternative~~ model structures contained an exponential function for snowmelt (M_e), and none of them included~~s~~ an exponential function for precipitation phase partition (ΔP_e). Most model structures ~~are~~ were the result of the combination of three to~~or~~ four ~~different~~ individual model structure modifications (seven ~~alternatives~~ model structures each). Four ~~alternatives~~ model structures contained two model modifications and two ~~alternatives~~ model structures contained~~include~~ five modifications. Perhaps most interestingly, two of the model ~~alternatives~~ structures ~~only~~ included only a single modification: an exponential snowmelt function, and a sine function for precipitation phase partition (ΔP_s). Nevertheless, these alternatives ~~have~~ had the lowest ranking amongst the selection. Overall, the top-ranking alternatives contained a seasonally varying degree-day factor ($C_{0,s}$) and an exponential snowmelt function, while other individual modifications resulted~~ed~~ in ~~more variable~~ more considerable model performance~~s~~ variability.

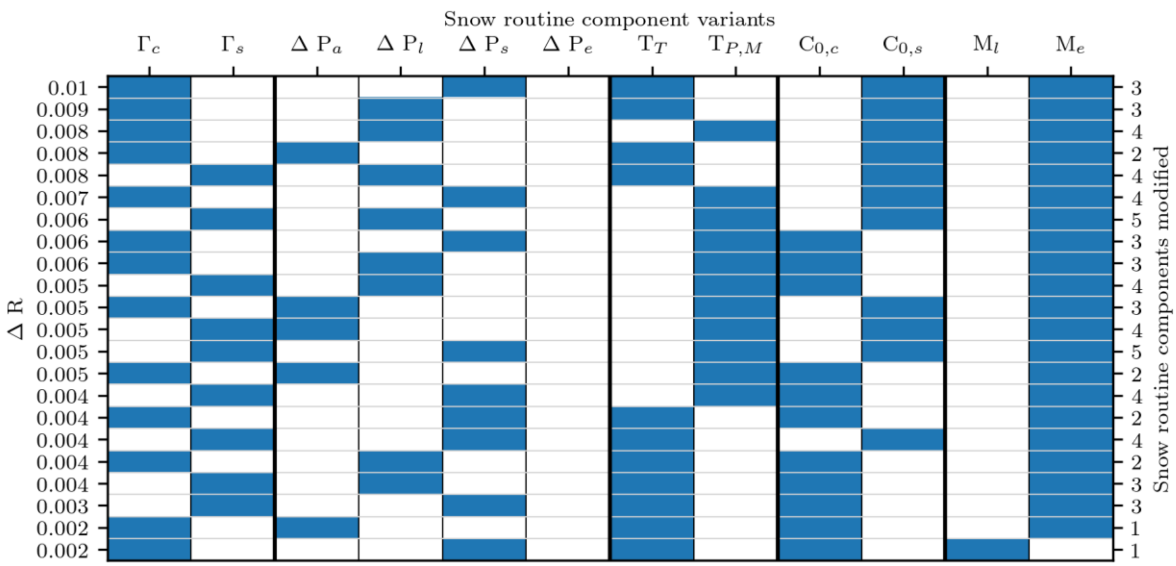


Figure 9 Ranked alternative structures to the snow routine of HBV that present positive relative ~~efficiency~~ model performance values with respect to HBV for model calibration and validation against snow water equivalent and stream runoff disaggregated by snow routine component variants. The alternatives for each of the considered model components are a linear and a seasonally-variable degree-day factor (*Γ_c* and *Γ_s* respectively); an abrupt, linear, sinusoidal, and exponential precipitation phase partition (*ΔP_a*, *ΔP_l*, *ΔP_s*, and *ΔP_e* respectively); a common and individualised threshold temperature for precipitation phase partition and snowmelt (*T_T* and *$T_{P,M}$* respectively); a constant and seasonally-variable degree-day factor (*$C_{0,c}$* and *$C_{0,s}$* respectively); and a linear and exponential (with no refreezing) melt function (*M_l* and *M_e* respectively). Every row contains one model structure with the selected variant for each of the components highlighted in blue. The median relative ~~efficiency~~ model performance for all modelling scenarios is given on the left y-axis while the number of model modifications in each alternative is provided in the right y-axis.

While not shown explicitly in this paper, the results obtained from the Monte Carlo sensitivity analysis showed that, even if some of the model structure variants (i.e. $T_{p,m}$, ΔP_e, or $C_{0,s}$) produced compensating effects on some of the model parameters (e.g. precipitation lapse rate, maximum storage in soil box, threshold for reduction of evaporation, or the shape coefficient), this effect was only observed for a reduced number of catchments. Most parameters showed no compensating effects at all. Overall, parameter values and their sensitivity tended to be reasonably consistent across all the tested model structure variants for most of the catchments in the study.

4 Discussion

~~From the results obtained in this study we can conclude that i~~It is ~~difficult~~ challenging to improve existing ~~hydrological~~ rainfall-runoff models and especially those that, like HBV, have successively been tested and applied in ~~proven capable of reproducing the hydrological behaviour~~ many~~of~~ catchments over a~~n extensive~~ range of environmental and geographical conditions (Bergström, 2006). ~~That being said~~Nevertheless, some of the proposed ~~modifications~~ alternative~~to the~~ snow routine ~~of the HBV~~ model structures that we tested in this study showed a generally positive impact on ~~the~~ model performance for simulating both snow water equivalent and stream runoff, albeit to different extents. We found that ~~T~~the most valuable modification to single HBV ~~model modifications~~ snow routine components for ~~modelling~~ rainfall-runoff modelling in mountainous catchments in Central Europe ~~are~~ was the use of an exponential snowmelt function and, to a lesser extent, a seasonally-varying degree-day factor. Another modification, using different thresholds for snowfall and snowmelt instead of a single threshold, ~~produces~~ produced a significant model performance improvement~~s~~ ~~towards~~regarding snow water equivalent, but ~~does~~ did not convey any advantage for simulating stream runoff.

We observed a significant differences in model performance changes between both objective functions when testing the different snow routine model structures.~~Continuing on the impact of model modifications on the different objective functions we considered in this study, we have seen that there are large differences between them.~~ Indeed, I~~i~~n general, the impact was~~is~~ more evident when simulating snow water equivalent than when~~, as the~~ simulat~~ing~~ed stream runoff, as the latter is the result of the combined model routines~~modulated and smoothened out by the other routines of the model~~ (i.e. snow, soil, groundwater, and routing routines), which partially compensate and mask any modifications made on the snow routine (Clark and Vrugt, 2006). Additionally, some of the modifications that improved the model performance ~~of the model to simulate~~against snow water equivalent, such as the use of an exponential function to define the solid and liquid phases of precipitation, ~~clearly penalise~~resulted in poorer the stream runoff simulations.

Unlike most modifications considered in this study, which are ~~simplifications~~ simple conceptualisations of complex processes, the use of an exponential function to describe precipitation phase partition and the use of a seasonally varying temperature lapse rate are both ~~modifications~~ formulations derived from empirical evidence (Magnusson et al., 2014; Rolland, 2002). Nevertheless, as we have previously discussed, neither of these modifications translate into an improvement of model performance for ~~any of the~~either objective functions. This might be ~~due to the fact that~~because, since ~~the~~ conceptual models such as HBV are based on simplifications and generalisations of the processes that occur in reality, ~~these~~ formulations based on accurate measurements of diverse~~the~~ processes do not align well with the other simplifications made in the model structure and/or behaviour at~~or~~ the chosen spatio-temporal resolution (Harder and Pomeroy, 2014; Magnusson et al., 2015).

Other modifications are relatively similar to each other, such as the case of using linear and sine functions to describe precipitation phase partition. Both these ~~conceptualisations~~ formulations require only one additional parameter, and perform almost identical~~. Indeed,~~ : the precipitation partition between rain and snow is exactly the same for both ~~conceptualisations~~ formulations for most of the transition temperature range except for~~. The only divergence is in~~ the tails, which are abrupt for the linear case and smooth for the sine one. Provided that the smooth transition is a more accurate description of the physical process which, in addition, avoids the introduction of discontinuities into the objective~~s~~ functions –~~,~~ which might complicate model calibration (Kavetski and Kuczera, 2007) –, and that both modifications include the same ~~degree of complexity~~ number of parameters and perform nearly identical, ~~we should favour~~ the most accurate description should be preferred. Nevertheless, some models, including HBV, continue to use the linear conceptualisation with the argument of simplicity.

Even if we did not observe differences in model performance as a function of different catchment characteristics, snow water equivalent tends to be underestimated in lowland catchments, while we observed no clear pattern for alpine catchments. Nevertheless, the limited number of high-elevation catchments in the dataset (only four of them have a mean elevation above

2000 m a.s.l.) combined with the generally small size, steep topography, relatively large glacierised areas, scarce vegetation and exposure to extreme weather conditions (such as strong wind gusts) of these catchments, makes it difficult to extract any relevant trends. That being said, in general, the different model structures tended to underestimate snow accumulation and delay the timing of the spring snowmelt season. Similar patterns have also been observed for other mountainous areas of Central Europe (Sleziak et al., 2020). We observed differences in model performance among the two geographical domains included in this study.

~~Most notably~~Among the different snow routine model structures, the modifications on precipitation phase partition penalised the model performance on most Czech catchments for simulating snow water equivalent, while having the opposite ~~impact~~ effect for Swiss catchments. The Czech catchments have a narrower elevation range compared to the Swiss catchments, in addition to an earlier and shorter snowmelt period. These characteristics may favour the simplification of an abrupt transition between rain and snow, while using gradual transitions between rain and snow might favour~~s~~ the more extended melt season and larger elevation ranges of the Swiss catchments. Another factor that may impact the results is the significant differences in model driving and validation data availability for each of the geographical domains (Günther et al., 2019; Meeks et al., 2017). Indeed, while in Czechia there ~~are~~ were a limited number of meteorological stations providing temperature, precipitation and snow water equivalent data, the Swiss catchments benefited from distributed data for the different catchment elevations, allowing for more accurate calibration of snow-related parameters. The difference in resolution between the Swiss and Czech input data might affect the obtained results, where each has its strengths and weaknesses. For instance, high-resolution data can become highly uncertain for individual grid cells while observational data may be affected by measurement errors and representativeness issues. Even so, the model performance variability of the different snow routine structures relative to the default HBV model should be similar for both cases.

~~Nevertheless~~Overall, even with the large differences in hydrological regime, catchment morphology, and data availability between the two geographical domains, the impact of the different ~~modifications~~ snow routine model structures ~~to~~ on model performance was in general comparable among them. ~~This is a relevant point because HBV aims to be a general-purpose hydrological model that is applicable to a range of geographical domains and both in areas with data wealth and scarcity. Achieving comparable model performances under these different conditions is an indication of the suitability and strength of the HBV model for such goals.~~

Regarding model complexity and uncertainty, we found that increasing the ~~complexity~~degree of detail, and thus the number of parameters of the model, generally translated into a broader range of model performance~~s~~ values, indicating that the uncertainty related to the model structure ~~was also~~ increased as well. This is a well-known problem of conceptual ~~hydrological~~ rainfall-runoff models, and the focus of many studies (Essery et al., 2013; Strasser et al., 2002). Additionally, we found that, for most cases, the median model performance increase with an increasing degree of detail~~model complexity~~ was not significant with respect to the performance range. This means that a more detailed~~n increasing~~ model ~~complexity~~ does not necessarily translate into better model performance, which is consistent with ~~other~~previous studies (Orth et al., 2015). This fact highlights the importance of carefully choosing the degree of ~~complexity~~ detail of the model based on the desired objectives to be achieved and the available data (Hock, 2003; Magnusson et al., 2015). Another important aspect is the uncertainty and robustness of the model's parameters. In conceptual models, such as the HBV model, model parameters can compensate for each other, which makes the interpretation of any model structure modifications rather challenging (Clark and Vrugt, 2006). Nevertheless, a Monte Carlo sensitivity analysis on the HBV model parameters showed that parameter values and sensitivity were consistent for most catchments and model structures in this study.

Even if an increased modeldegree of detail in the processes description complexity is not desirable by itself, as it can lead to overparameterisation and equifinality issues, if implemented in a sensible way, it can also improve the performance of hydrological rainfall-runoff models if it responds to specific needs and/or available data, among others. Indeed, 22 of the 63 model structure alternatives that we tested in this study (all of them conveying an increase of model complexitydetail, and therefore an increase in the number of parameters) convey and increase ofin model performance with respect to HBV for both model calibration and simulation against both objective functionsall cases. Furthermore, out of these 22 alternatives, only two of them consist of a single model structure modification, while most have 3 or 4 modifications. Nevertheless, all of these alternatives share common traits, such as using an exponential snowmelt function with no refreezing. Indeed, aAlmost all model structures that do not have this particular modification perform worse than HBV in at least one simulation scenario.

Considering these results, iIt is reasonable to state that, while the increased realism degree of detail arising from the interplay among the different model structure modifications in these model structure alternatives play a role in improving the model performance, this is mainly the result of a few dominant modifications. This way, the use of an exponential snowmelt function is the most valuable single modification with a median performance increase of 0.002 with respect to HBV for all simulations (with individual performance increases over 0.1). However, when we combine it with a seasonal degree-day factor, we achieve a median performance increase of 0.008, almost the highest performance increase amongst all model alternatives. Adding further complexity detail to the model does not convey significant improvements with respect tofor this model structure. Consequently, if we were to implement any modifications to the model, they would be to substitute the linear snowmelt and refreezing conceptualisation by an exponential snowmelt function, and replacing the constant degree-day factor by a seasonally varying one, in that order.

Finally, it is important to mention that these results are only valid for the selected study areas and cannot be extrapolated to all the different alpine and snow-covered areas regions around the world as. Tthe different processes involved in different geologic, geographic, climatological, and hydrological settings are likely to favour different conceptualisations formulations of the snow processes. So, while the proposed modifications might be useful to tailor the model to Central-European alpine catchments, we still consider the default model structure of HBV as a very capable general purpose model that can be used in a variety of settings and for different purposes.

5 Conclusions

We evaluated the suitability of different temperature-based snow routine model structures for rainfall-runoff modelling in Alpine areas of Central Europe. More specifically performance of the snow routine of the HBV model and explored possible modifications to make it more realistic and relevant for hydrological modelling in Alpine areas in Central Europe. Wwe tested a number of modifications to all of the differenteach of the components of the snow routine of the HBV model over a large number of catchments covering a range of geographical settings, and different data availability conditions based on their ability to reproduce both snow water equivalent and stream runoff. We found that the results differ greatly across the different catchments, objective functions, and simulation type (i.e. calibration/validation). However, the resultsStill they allow drawing the following general conclusions regarding the value of the different model modifications for the performance in mountainous catchments in Central Europesnow routine model structures:

- The comparatively simple default structure of the HBV model performs well whenfor simulating snow-related processes and their impact to stream runoff in most of the examined catchments.
- Specific modifications to the formulation of certain processes in the snow routine structure of the model improve the performance of the model for estimating snow processes and, to a lesser extent, for simulating stream runoff.
- An exponential snowmelt function with no refreezing is the single most valuable overall modification to the snow routine structure of HBV, followed by a seasonally variable degree-day factor.
- Adding further detail to the snow routine model structure n increased model complexity does not, by itself, add any value to on the ability of the model to reproduce snow water equivalent and nor stream runoff. A careful examination of the design choices of the model for the given application, data availability, and objective purpose – such as the one presented in this study – is crucial to ensure that the model conceptualisation is suitable and to provide guidance on potential model improvementsis a significantly better approach.
- An exponential snowmelt function with no refreezing is the single most valuable overall modification to the snow routine structure, followed by a seasonally variable degree-day factor. Adding further complexity to the model does not necessarily translate in significant performance improvements.

Thesе specific results obtained in this study are not transferrable to other geographical domains, models, or purposes., butNevertheless, the methodology presented here is relevant to the general degree-day approach. It may, therefore, be used to assess the suitability of model design choices in temperature-based snow routines in other rainfall-runoff models in different circumstances.

Appendix A: HBV model performance

Table A1 HBV median calibration and validation performance values – both for snow water equivalent, R_W, and logarithmic stream runoff, R_{lnQ} – for each catchment and analysis period.

ID	Period 1				Period 2			
	R_W		R_{lnQ}		R_W		R_{lnQ}	
	Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation
CH-101	0.75	0.68	0.92	0.87	0.82	0.73	0.87	0.81
CH-102	0.59	0.5	0.85	0.8	0.8	0.61	0.87	0.83
CH-103	0.74	0.57	0.91	0.89	0.84	0.69	0.86	0.84
CH-104	0.77	0.69	0.83	0.76	0.81	0.79	0.75	0.63
CH-105	0.75	0.73	0.78	0.75	0.8	0.78	0.74	0.69
CH-106	0.77	0.65	0.85	0.82	0.84	0.69	0.79	0.75
CH-107	0.81	0.78	0.85	0.82	0.81	0.78	0.83	0.8
CH-108	0.84	0.81	0.87	0.85	0.84	0.75	0.83	0.81
CH-109	0.84	0.79	0.88	0.84	0.81	0.73	0.84	0.77
CH-110	0.92	0.9	0.87	0.85	0.91	0.86	0.84	0.82
CH-111	0.8	0.79	0.89	0.88	0.8	0.78	0.84	0.81
CH-112	0.81	0.74	0.87	0.85	0.82	0.74	0.82	0.79
CH-113	0.92	0.91	0.86	0.85	0.88	0.86	0.85	0.84
CH-201	0.8	0.81	0.81	0.78	0.78	0.74	0.74	0.69
CH-202	0.92	0.86	0.89	0.86	0.92	0.87	0.85	0.8
CH-203	0.88	0.87	0.9	0.86	0.84	0.83	0.9	0.87
CH-204	0.9	0.87	0.88	0.72	0.85	0.82	0.83	0.56
CH-205	0.92	0.91	0.85	0.82	0.9	0.89	0.89	0.86
CH-206	0.9	0.87	0.94	0.92	0.85	0.82	0.95	0.92
CH-301	0.78	0.69	0.94	0.92	0.91	0.84	0.95	0.94
CH-302	0.7	0.62	0.94	0.92	0.87	0.83	0.94	0.93
CH-303	0.77	0.71	0.92	0.7	0.83	0.64	0.91	0.86
CZ-101	0.86	0.77	0.84	0.81	0.92	0.79	0.82	0.65
CZ-102	0.86	0.76	0.88	0.82	0.92	0.78	0.85	0.74
CZ-103	0.86	0.77	0.85	0.8	0.92	0.79	0.83	0.73
CZ-104	0.91	0.88	0.81	0.74	0.75	0.72	0.83	0.72
CZ-105	0.9	0.8	0.81	0.76	0.94	0.87	0.8	0.73
CZ-106	0.85	0.67	0.84	0.8	0.93	0.88	0.81	0.72
CZ-107	0.89	0.84	0.8	0.74	0.92	0.87	0.87	0.77
CZ-201	0.92	0.82	0.81	0.76	0.91	0.85	0.7	0.64
CZ-202	0.92	0.83	0.69	0.47	0.91	0.85	0.66	0.44

CZ-203	0.92	0.82	0.84	0.59	0.91	0.85	0.84	0.68
CZ-204	0.92	0.83	0.74	0.71	0.91	0.85	0.89	0.87
CZ-205	0.92	0.83	0.88	0.84	0.91	0.86	0.81	0.77
CZ-206	0.86	0.63	0.83	0.79	0.89	0.78	0.81	0.76
CZ-207	0.87	0.62	0.79	0.58	0.89	0.78	0.81	0.65
CZ-208	0.87	0.62	0.66	0.55	0.89	0.79	0.83	0.73
CZ-209	0.95	0.92	0.79	0.7	0.93	0.88	0.82	0.77
CZ-210	0.95	0.92	0.81	0.74	0.93	0.89	0.81	0.6
CZ-211	0.87	0.62	0.83	0.8	0.89	0.78	0.85	0.75
CZ-212	0.88	0.62	0.79	0.74	0.9	0.8	0.83	0.76
CZ-213	0.87	0.64	0.81	0.77	0.9	0.79	0.82	0.8
CZ-301	0.93	0.87	0.81	0.7	0.97	0.92	0.86	0.78
CZ-302	0.93	0.87	0.84	0.71	0.97	0.92	0.85	0.67
CZ-303	0.94	0.87	0.86	0.8	0.97	0.91	0.87	0.77
CZ-304	0.93	0.87	0.89	0.86	0.97	0.92	0.87	0.82
CZ-305	0.79	0.67	0.82	0.7	0.81	0.75	0.83	0.74
CZ-401	0.95	0.89	0.84	0.81	0.9	0.85	0.78	0.69
CZ-402	0.92	0.9	0.74	0.7	0.93	0.92	0.82	0.77
CZ-403	0.92	0.89	0.81	0.78	0.96	0.94	0.86	0.82
CZ-404	0.91	0.89	0.75	0.73	0.96	0.94	0.82	0.8
CZ-405	0.91	0.89	0.79	0.76	0.96	0.93	0.87	0.84
CZ-406	0.85	0.82	0.78	0.71	0.92	0.86	0.8	0.66
CZ-407	0.85	0.82	0.73	0.67	0.92	0.87	0.74	0.7
Median	0.86	0.78	0.83	0.77	0.89	0.82	0.83	0.76

6 Data availability

Meteorological and hydrological data to calibrate the HBV model were obtained from the Swiss Federal Office of Meteorology and Climatology, the Swiss Federal Office for the Environment, and the Czech Hydrometeorological Institute. The HBV model outputs are available from the first author upon request.

7 Author contribution

JS initiated the study. MGL developed the methodology and performed all analyses. MGL, NG, and MJ prepared the input meteorological and hydrological data used to calibrate the HBV model. MJPV performed all the necessary modifications to the source code of the HBV model. MGL prepared the manuscript with contributions from all co-authors.

8 Competing interests

The authors declare that they have no conflict of interest.

9 Acknowledgements

This project was partially funded by the Swiss Federal Office for the Environment (FOEN). The contribution of Michal Jenicek was supported by the Czech Science Foundation, project no. GA18-06217Y. We thank reviewers Juraj Parajka, Valentina Premier, María José Polo, and Thomas Skaugen for the valuable comments and suggestions that greatly improved this paper.

References

Avanzi, F., De Michele, C., Morin, S., Carmagnola, C. M., Ghezzi, A. and Lejeune, Y.: Model complexity and data requirements in snow hydrology: seeking a balance in practical applications, Hydrol. Process., 30(13), 2106–2118, doi:10.1002/hyp.10782, 2016.

Barnett, T. P., Adam, J. C. and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, Nature, 438(7066), 303–309, doi:10.1038/nature04141, 2005.

Berghuijs, W. R., Woods, R. A. and Hrachowitz, M.: A precipitation shift from snow towards rain leads to a decrease in streamflow-supplement, Nat. Clim. Chang., 4(7), 583–586, doi:10.1038/NCLIMATE2246, 2014.

Bergström, S.: The HBV Model, in Computer Models of Watershed Hydrology, edited by V. P. Singh, pp. 443–476, Water Resources Publications, Highlands Ranch, CO., 1995.

Bergström, S.: Experience from applications of the HBV hydrological model from the perspective of prediction in ungauged basins, IAHS Publ., 307, 97–107, 2006.

Beven, K. J.: Environmental Modelling: An Uncertain Future?, Routledge, London., 2008.

Braun, J. N. and Renner, C. B.: Application of a conceptual runoff model in different physiographic regions of Switzerland, Hydrol. Sci. J., 37(3), 217–231, doi:10.1080/02626669209492583, 1992.

Clark, M. P. and Vrugt, J. A.: Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters, Geophys. Res. Lett., 33(6), 1–5, doi:10.1029/2005GL025604, 2006.

Collins, E. H.: Relationship of Degree-Days above Freezing to Runoff, Eos, Trans. Am. Geophys. Union, 15(2), 624–629,

1934.
Dai, A.: Temperature and pressure dependence of the rain-snow phase transition over land and ocean, Geophys. Res. Lett., 35(12), 1–7, doi:10.1029/2008GL033295, 2008.
Debele, B., Srinivasan, R. and Gosain, A. K.: Comparison of process-based and temperature-index snowmelt modeling in SWAT, Water Resour. Manag., 24(6), 1065–1088, doi:10.1007/s11269-009-9486-2, 2010.
Dietz, A. J., Kuenzer, C., Gessner, U. and Dech, S.: Remote sensing of snow - a review of available methods, Int. J. Remote Sens., 33(13), 4094–4134, doi:10.1080/01431161.2011.640964, 2012.
Essery, R., Morin, S., Lejeune, Y. and Ménard, C. B.: A comparison of 1701 snow models using observations from an alpine site, Adv. Water Resour., 55, 131–148, doi:10.1016/j.advwatres.2012.07.013, 2013.
Fang, S., Xu, L., Pei, H., Liu, Y., Liu, Z., Zhu, Y., Yan, J. and Zhang, H.: An Integrated Approach to Snowmelt Flood Forecasting in Water Resource Management, IEEE Trans. Ind. Informatics, 10(1), 548–558, doi:10.1109/TII.2013.2257807, 2014.
Flügel, W.-A.: Hydrological response units (HRUs) to preserve basin heterogeneity in hydrological modelling using PRMS/MMS - case study in the Bröl basin, Germany, in Modelling and Management of Sustainable Basin-scale Water Resource Systems, vol. 231, pp. 79–87, International Association of Hydrological Sciences (IAHS), Boulder, CO., 1995.
FOEN: Observed Discharge Time Series of Swiss Rivers, 2017.
Frei, C.: Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances, Int. J. Climatol., 34, 1585–1605, doi:10.1002/joc.3786, 2014.
Frei, C. and Schär, C.: A precipitation climatology of the Alps from high-resolution rain-gauge observations, Int. J. Climatol., 18, 873–900, doi:10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9, 1998.
Frei, C., Schöll, R., Fukutome, S., Schmidli, J. and Vidale, P. L.: Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, J. Geophys. Res., 111, doi:10.1029/2005JD005965, 2006.
Freudiger, D., Kohn, I., Seibert, J., Stahl, K. and Weiler, M.: Snow redistribution for the hydrological modeling of alpine catchments, Wiley Interdiscip. Rev. Water, 4(5), e1232, doi:10.1002/wat2.1232, 2017.
Frey, S. and Holzmann, H.: A conceptual, distributed snow redistribution model, Hydrol. Earth Syst. Sci., 19(11), 4517–4530, doi:10.5194/hess-19-4517-2015, 2015.
Froidurot, S., Zin, I., Hingray, B. and Gautheron, A.: Sensitivity of Precipitation Phase over the Swiss Alps to Different Meteorological Variables, J. Hydrometeorol., 15(2), 685–696, doi:doi:10.1175/JHM-D-13-073.1, 2014.
Gabbi, J., Carenzo, M., Pellicciotti, F., Bauder, A. and Funk, M.: A comparison of empirical and physically based glacier surface melt models for long-term simulations of glacier response, J. Glaciol., 60(224), 1199–1207, doi:10.3189/2014JoG14J011, 2014.
Griessinger, N., Seibert, J., Magnusson, J. and Jonas, T.: Assessing the benefit of snow data assimilation for runoff modeling in Alpine catchments, Hydrol. Earth Syst. Sci., 20(9), 3895–3905, doi:10.5194/hess-20-3895-2016, 2016.
Günther, D., Marke, T., Essery, R. and Strasser, U.: Uncertainties in Snowpack Simulations—Assessing the Impact of Model

Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance, Water Resour. Res., 55(4), 2779–2800, doi:10.1029/2018WR023403, 2019.

Hannah, D. M., Brown, L. E., Milner, A. M., Gurnell, A. M., McGregor, G. R., Petts, G. E., Smith, B. P. G. and Snook, D. L.: Integrating climate-hydrology-ecology for alpine river systems, Aquat. Conserv. Mar. Freshw. Ecosyst., 17, 636–656, doi:10.1002/aqc.800, 2007.

Harder, P. and Pomeroy, J. W.: Hydrological model uncertainty due to precipitation-phase partitioning methods, Hydrol. Process., 28(14), 4311–4327, doi:10.1002/hyp.10214, 2014.

Harpold, A. A., Kaplan, M. L., Zion Klos, P., Link, T., McNamara, J. P., Rajagopal, S., Schumer, R. and Steele, C. M.: Rain or snow: Hydrologic processes, observations, prediction, and research needs, Hydrol. Earth Syst. Sci., 21(1), 1–22, doi:10.5194/hess-21-1-2017, 2017.

He, Z. H., Parajka, J., Tian, F. Q. and Blöschl, G.: Estimating degree-day factors from MODIS for snowmelt runoff modeling, Hydrol. Earth Syst. Sci., 18(12), 4773–4789, doi:10.5194/hess-18-4773-2014, 2014.

Helbig, N., Van Herwijnen, A., Magnusson, J. and Jonas, T.: Fractional snow-covered area parameterization over complex topography, Hydrol. Earth Syst. Sci., 19(3), 1339–1351, doi:10.5194/hess-19-1339-2015, 2015.

Herrero, J. and Polo, M. J.: Evaposublimation from the snow in the Mediterranean mountains of Sierra Nevada (Spain), Cryosphere, 10(6), 2981–2998, doi:10.5194/tc-10-2981-2016, 2016.

Hock, R.: A distributed temperature-index ice- and snowmelt model including potential direct solar radiation, J. Glaciol., 45(149), 101–111, doi:10.1017/S0022143000003087, 1999.

Hock, R.: Temperature index melt modelling in mountain areas, J. Hydrol., 282(1–4), 104–115, doi:10.1016/S0022-1694(03)00257-9, 2003.

Hock, R.: Glacier melt: a review of processes and their modelling, Prog. Phys. Geogr., 29(3), 362–391, doi:10.1191/0309133305pp453ra, 2005.

Hottelet, C., Blažková, Š. and Bičík, M.: Application of the ETH Snow Model to Three Basins of Different Character in Central Europe, Nord. Hydrol., 25(1–2), 113–128, 1994.

Jamieson, B. and Stethem, C.: Snow avalanche hazards and management in Canada: Challenges and progress, Nat. Hazards, 26(1), 35–53, doi:10.1023/A:1015212626232, 2002.

Jenicek, M. and Ledvinka, O.: Importance of snowmelt contribution to seasonal runoff and summer low flows in Czechia, Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2019-611, 2020.

Kane, D. L., Gieck, R. E. and Hinzman, L. D.: Snowmelt Modeling at Small Alaskan Arctic Watershed, J. Hydrol. Eng., 2(4), 204–210, doi:10.1061/(ASCE)1084-0699(1997)2:4(204), 1997.

Kavetski, D. and Kuczera, G.: Model smoothing strategies to remove microscale discontinuities and spurious secondary optima im objective functions in hydrological calibration, Water Resour. Res., 43(3), 1–9, doi:10.1029/2006WR005195, 2007.

Lindström, G., Johansson, B., Persson, M., Gardelin, M. and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, J. Hydrol., 201, 272–288, doi:10.1016/S0022-1694(97)00041-3, 1997.

Magnusson, J., Gustafsson, D., Hüsler, F. and Jonas, T.: Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods, Water Resour. Res., 50, 7816–7835, doi:10.1002/ 2014WR015302, 2014.

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A. and Jonas, T.: Evaluating snow models with varying process representations for hydrological applications, Water Resour. Res., 51(4), 2707–2723, doi:10.1002/2014WR016498, 2015.

Martinec, J.: The degree–day factor for snowmelt runoff forecasting, in IUGG General Assembly of Helsinki, vol. 51, pp. 468–477, IAHS Commission of Surface Waters, Helsinki, Finland., 1960.

Matsuo, T. and Sasyo, Y.: Melting of Snowflakes below Freezing Level in the Atmosphere, J. Meteorol. Soc. Japan, 59(1), 10–25, doi:10.2151/jmsj1965.59.1_10, 1981.

Meeks, J., Moeck, C., Brunner, P. and Hunkeler, D.: Infiltration under snow cover: Modeling approaches and predictive uncertainty, J. Hydrol., 546, 16–27, doi:10.1016/j.jhydrol.2016.12.042, 2017.

Merz, R. and Blöschl, G.: A process typology of regional floods, Water Resour. Res., 39(12), 1–20, doi:10.1029/2002WR001952, 2003.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I: A discussion of principles, J. Hydrol., 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Ohmura, A.: Physical Basis for the Temperature-Based Melt-Index Method, J. Appl. Meteorol., 40(4), 753–761, doi:10.1175/1520-0450(2001)040<0753:PBFTTB>2.0.CO;2, 2001.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J. and Zappa, M.: Does model performance improve with complexity? A case study with three hydrological models, J. Hydrol., 523, 147–159, doi:10.1016/j.jhydrol.2015.01.044, 2015.

Pellicciotti, F., Brock, B., Strasser, U., Burlando, P., Funk, M. and Corripio, J.: An enhanced temperature-index glacier melt model including the shortwave radiation balance: development and testing for Haut Glacier d'Arolla, Switzerland, J. Glaciol., 51, doi:10.3189/172756505781829124, 2005.

Pomeroy, J. W. and Marks, D., Eds.: Hydrometeorlogical data from mountain and alpine research catchments [Special issue], in Earth System Science Data., 2020.

Quick, M. C. and Pipes, A.: U.B.C. watershed model, Hydrol. Sci. Bull., 22(1), 153–161, doi:10.1080/02626667709491701, 1977.

Riboust, P., Thirel, G., Moine, N. Le and Ribstein, P.: Revisiting a simple degree-day model for integrating satellite data: implementation of swe-sca hystereses, J. Hydrol. Hydromechanics, 67(1), 70–81, doi:10.2478/johh-2018-0004, 2019.

Rico, M., Benito, G., Salgueiro, A. R., Díez-Herrero, A. and Pereira, H. G.: Reported tailings dam failures. A review of the European incidents in the worldwide context, J. Hazard. Mater., 152(2), 846–852, doi:10.1016/j.jhazmat.2007.07.050, 2008.

Rolland, C.: Spatial and Seasonal Variations of Air Temperature Lapse Rates in Alpine Regions, J. Clim., 16, 1032–1046, doi:10.1175/1520-0442(2003)016<1032:SASVOA>2.0.CO;2, 2002.

Schaefli, B. and Gupta, H. V.: Do Nash values have value?, Hydrol. Process., 21, 2075–2080, doi:10.1002/hyp.6825, 2007.

Schweizer, J., Jamieson, J. B. and Schneebeli, M.: Snow avalanche formation, Rev. Geophys., 41(4), doi:10.1029/2002RG000123, 2003.

Seibert, J.: Regionalisation of parameters for a conceptual rainfall-runoff model, Agric. For. Meteorol., 98–99, 279–293, doi:10.1016/S0168-1923(99)00105-7, 1999.

Seibert, J.: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, Hydrol. Earth Syst. Sci., 4(5), 215–224, 2000.

Seibert, J. and Vis, M. J. P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrol. Earth Syst. Sci., 16, 3315–3325, doi:10.5194/hess-16-3315-2012, 2012.

Sicart, J. E., Hock, R. and Six, D.: Glacier melt, air temperature, and energy balance in different climates: The Bolivian Tropics, the French Alps, and northern Sweden, J. Geophys. Res. Atmos., 113(24), 1–11, doi:10.1029/2008JD010406, 2008.

Sims, E. M. and Liu, G.: A Parameterization of the Probability of Snow–Rain Transition, J. Hydrometeorol., 16(4), 1466–1477, doi:10.1175/JHM-D-14-0211.1, 2015.

Skaugen, T., Luijting, H., Saloranta, T., Vikhamar-Schuler, D. and Müller, K.: In search of operational snow model structures for the future – comparing four snow models for 17 catchments in Norway, Hydrol. Res., 49(6), 1929–1945, doi:10.2166/nh.2018.198, 2018.

Sleziak, P., Szolgay, J., Hlavčová, K., Danko, M. and Parajka, J.: The effect of the snow weighting on the temporal stability of hydrologic model efficiency and parameters, J. Hydrol., 583(September 2019), doi:10.1016/j.jhydrol.2020.124639, 2020.

Strasser, U., Etchevers, P. and Lejeune, Y.: Inter-comparison of two snow models with different complexity using data from an alpine site, Nord. Hydrol., 33(1), 15–26, doi:10.2166/nh.2002.0002, 2002.

Sui, J. and Koehler, G.: Rain-on-snow induced flood events in southern Germany, J. Hydrol., 252(1–4), 205–220, doi:10.1016/S0022-1694(01)00460-7, 2001.

Uhlenbrook, S., Seibert, J., Leibundgut, C. and Rodhe, A.: Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure, Hydrol. Sci. J., 44(5), 779–797, doi:10.1080/02626669909492273, 1999.

Valéry, A., Andréassian, V. and Perrin, C.: “As simple as possible but not simpler”: What is useful in a temperature-based snow-accounting routine? Part 1 - Comparison of six snow accounting routines on 380 catchments, J. Hydrol., 517, 1166–1175, doi:10.1016/j.jhydrol.2014.04.059, 2014.

Verdhen, A., Chahar, B. R. and Sharma, O. P.: Snowmelt modelling approaches in watershed models: Computation and comparison of efficiencies under varying climatic conditions, Water Resour. Manag., 28(11), 3439–3453, doi:10.1007/s11269-014-0662-7, 2014.

Vis, M., Knight, R., Pool, S., Wolfe, W. and Seibert, J.: Model calibration criteria for estimating ecological flow characteristics, Water, 7(5), 2358–2381, doi:10.3390/w7052358, 2015.

Viviroli, D., Gurtz, J. and Zappa, M.: The hydrological modelling system PREVAH, edited by S. Steiner, Geographica Bernensia P40, Berne, Switzerland., 2007.

Walter, M. T., Brooks, E. S., McCool, D. K., King, L. G., Molnau, M. and Boll, J.: Process-based snowmelt modeling: does it require more input data than temperature-index modeling?, J. Hydrol., 300(1–4), 65–75, doi:10.1016/j.jhydrol.2004.05.002, 2005.

Watson, B. M. and Putz, G.: Comparison of Temperature-Index Snowmelt Models for Use within an Operational Water Quality Model, J. Environ. Qual., 43(1), 199–207, doi:10.2134/jeq2011.0369, 2014.

Weingartner, R. and Aschwanden, H.: Discharge Regime - the Basis for the Estimation of Average Flows, Plate 5.2, in Hydrological Atlas of Switzerland, edited by R. Weingartner and S. M., Swiss Federal Office for Water and Geology, Bern, Switzerland., 1989.

Wever, N., Fierz, C., Mitterer, C., Hirashima, H. and Lehning, M.: Solving Richards Equation for snow improves snowpack meltwater runoff estimations in detailed multi-layer snowpack model, Cryosphere, 8(1), 257–274, doi:10.5194/tc-8-257-2014, 2014.

Zuzel, J. F. and Cox, L. M.: Relative importance of meteorological variables in snowmelt, Water Resour. Res., 11(1), 174–176, doi:10.1029/WR011i001p00174, 1975.