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

Black text: Reviewer comment

Blue text: Authors' response

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 attach some figures resulting from these analyses at the end of the comment (a caption is 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. It is also important to note that, even if some parameter profiles might look quite flat in these figures, hinting to equifinality issues, in most cases this is an artefact form forcing the same yaxis range across all subplots, which makes some of the parameter shapes difficult to appreciate. As we previously mentioned, we took consideration of both potential equifinality and parameter compensation issues in our analysis. We will emphasise this in the revised manuscript.

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.

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 will clarify this in the revision.

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. That being said, we will revise the manuscript to ensure that the terminology we use is appropriate and concise, and we will rewrite 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, I.476, p.25, I 520-25, p.26, I563). "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 will revise the passages mentioned by the reviewer to ensure that the demands for objectivity are 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 will, of course, consider this comment when revising the manuscript and we will replace the generic references to "hydrology" by more concise terms. Additionally, we will specify that we do not investigate processes other than snowpack accumulation and evolution and its impact on stream runoff (i.e. no subsurface, chemical... processes).

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. Nevertheless, we understand the concerns of the reviewer, and we will, therefore, add an appendix with the default HBV model performance values for all catchments, objective functions, and calibration/validation efforts included in this study. We believe that, based

on this table and on the figures included in the manuscript, the reader will have all the necessary information to evaluate this contribution. Additionally, if the editor and reviewer find this useful, we could provide detailed figures similar to Figure 4 (but including model validation results) for all catchments in the study in another appendix (54 figures). We provide these figures at the end of this response for the reviewer to evaluate (caption only provided for Figure 55).

Specific comments

P1, I.18, "popular" subjective

We agree with the reviewer that this is a subjective expression. We will rephrase the text to emphasise that this model has been (and still is) widely used in many different settings.

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

We will rephrase 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 will rephrase the sentence for clarification.

P2, I.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 will rephrase this sentence to make it fit better in the paragraph.

P2, I.52 ..available at..

We will correct this mistake.

P2, I.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 will address this by explicitly mentioning lumped and semi-distributed models that are based on these approaches. We thank the reviewer for providing this reference.

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

We intended to say "bucket-type model" / "conceptual model". We will correct this error in the revision.

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

We refer to the reply to major comment 1 and 2 above. We will revise the text to ensure that the terminology we used is concise and relevant.

P4, I.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, that 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 will rephrase 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 will reformulate 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. Nevertheless, we understand the confusion of the reviewer, and we will, therefore, add it to the table.

Regarding the calibration parameters, we argue that including all calibration parameters is out of the scope of this table. We will instead add 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, I.131. Heading, "Temperature and precipitation lapse rates"

We will modify the section header (see also the previous comment).

P7, I.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, I.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 subdiurnal 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 will rephrase this sentence to be more concise in the terminology.

P12, I.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, I.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, I.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 will revise the manuscript to ensure that the terminology is concise and relevant throughout the text.

P14, I.322..performance for...

We will correct the error.

P14, I.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, I.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 will rephrase 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.

P20, I.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_I , ΔP_s or ΔP_e). We will rephrase the paragraph to make it easier to understand.

P20, I.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 and will change the table accordingly.

P23, I.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 will rephrase this to make it more concise.

P24, I.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 overparameterization, 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. In the discussion, we will delve deeper into the reasons why such models are difficult to improve.

P24, I.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 will rephrase the sentence taking the previous discussion into account.

P25, I.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 will rephrase the sentence to make it clearer.

P26, I.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 will make this sentence more specific.

P26, I.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 will include in the revision. We hope that the aforementioned changes provide enough evidence to support this conclusion.

P26, I.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.

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, <u>https://doi.org/10.2166/nh.2018.198</u>

Sensitivity analysis



Figure 1. Sensitivity analysis on parameters other than those from the snow routine for a catchment CH-101. Each subplot shows the model performance (y-axis) as a function of the values of a single free parameter (x-axis) with all other parameters set at the median best value of all 10 calibration trials for each of the single modifications to the snow routine of the model (see Table 1 in the manuscript). These results shown here are for model calibration on period 1.





















0.6 LP

0.8

0.4

1.0

1

3 4 BETA

4

5 6

2

 ${\rm R}_{\rm ln(Q)}$ 0.40.2

100 200 300 FC

400 500





















CZ-107



CZ-202









0.6 LP

0.8

0.4

1.0

1

3 BETA

4

5 6

2

0.20.0

100 200 300 FC

400 500







CZ-212







0.6 LP

0.8

1.0

1

0.4

3 4 BETA

4

5 6

2

0.20.0

100 200 300 FC

400 500















Hydrographs



Figure 55 Time series (October 2003 – September 2004) for the catchment CH-101. Top: daily mean air temperature and total precipitation. Middle: model calibration results. Bottom: model validation results. The model calibration and validation are further subdivided into (top) catchment-average observed (grey line) and simulated snow water equivalent (HBV in blue and the model structure modification including a seasonally-varying degree-day factor, CO,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.






































































































