

Reviewer comments are in blue, and responses are in black.

Firstly, we would like to thank Anonymous Reviewer 1 for the positive and constructive comments.

We provide this initial response to open the discussion and outline the changes we plan to make when we revise the manuscript. We will wait for the additional review(s) to become available before actually implementing these changes, however.

This paper presents an uncertainty analysis of snow simulations with the model WaSiM in a mountainous region in western Switzerland using multi-objective pre- and post-calibration methods including the usage of Landsat 8 images. The paper is well-written and the authors took especially great care in producing good and understandable figures as well as supplementary material. Moreover, it seems that the authors have a comprehensive overview on literature close to their study.

We are grateful that the efforts made to present a well-written paper with sufficient background context, appropriate figures, and extensive supplementary material were appreciated by the reviewer.

However, on the one side as mentioned by the authors themselves, the absence of ‘real’ continuous snow measurements within the study site is a clear limitation of this study, but on the other side the authors try the best in including ‘reconstructed snow measurements’.

The concern about ‘real’ continuous snow measurements may be a semantic issue. To be clear, we used continuous measurements of snow depth at two stations which were located in the vicinity of the target catchments. The simulation domain was extended in order to include these locations. In our view, these measurements can be described as ‘real’ continuous snow measurements.

To enable more meaningful comparisons to be made between our simulations and the available (time-series) observations, measured snow depths were converted into Snow Water Equivalent (SWE) via a modelling process. This process was different at each station according to data availability. Other studies using WaSiM did not attempt this extra step, and simply presented comparisons between simulated SWE and measured snow depth (e.g. Warscher et al., 2013).

In fact, some continuous snow depth measurements were attempted at the internal stations, but generally yielded gap filled and quite uncertain data that were not considered suitable for inclusion in our calibration process. That said, we certainly agree that more extensive in situ measurements – especially repeated SWE surveys at different locations in the study area – would have been extremely valuable. Unfortunately, the steep and remote (no ski installations etc.) nature of the study catchments mean that avalanche risk can be considerable yet uncertain, and so fieldwork in winter and spring is strongly limited. Faced with these challenges, as the reviewer highlights, we tried our best to develop datasets that would be as informative as possible in constraining our model.

Incidentally, a possibility for future work would be to undertake LiDAR scanning under summer and then winter conditions (e.g. Cochand et al., 2019) in order to generate (a) high resolution snapshot(s) of snow depths. Of course, a helicopter flight would be required to generate a catchment scale map, which brings cost implications, and furthermore a model to predict density would still be required to convert to water equivalent.

I have some general and some specific points, which should be considered carefully before publication.

General:

The results should also be separated in snow accumulation and snow ablation phase (as you shortly mentioned the two phases on p.38, l.819 and in the abstract). I guess there would be then distinct differences in your calibration. This should be presented and discussed.

As mentioned in the abstract, it is true that both snow accumulation and ablation need to be accurately simulated if reliable patterns of meltwater arrival at the land surface are ultimately to be generated. Clearly, some of the parameters subjected to calibration relate only to one of these two phases. For instance, the undercatch correction factors purely affect accumulation, whilst the longwave correction parameter affects ablation. However, other parameters are shared between these process components (e.g. rain snow threshold temperature, which also dictates whether melt can occur if the energy balance is favourable) or neither of them (e.g. the parameter concerned with the redistribution module). Additionally, the moderate elevation of the catchment meaning that parts of it can be affected by repeated cycles of accumulation and melt within a single winter season. This meant that even dividing the simulated period up into distinct accumulation and melt periods would not have been straightforward. As a result of these two facts, we elected not to separately calibrate the model for the accumulation and ablation phases, but rather sought parameters that gave the best overall representation of the dynamics according to the constructed objective function.

A related issue that we could have dealt with in a more sophisticated fashion is that fact that the catchment usually becomes snow covered very quickly in winter, whereas the melt in summer leads to a much more gradual evolution of snow extents. Therefore, during the onset of winter conditions, small errors in simulated snow onset (either slightly too early or too later, perhaps due to an insufficient air temperature measurements during a given period and attendant uncertainties in the interpolated grids) would lead to very large errors on these spatial statistics, but these would most probably have only relatively limited hydrological importance (due to the lower snowpack water storage early in the season). In contrast, any large errors in the simulated snow extent in spring would likely have much more significant hydrological implications. In theory, this issue could have been partially addressed by weighting the observed snow extent maps that correspond to the “end” of the season more highly than those corresponding to the beginning, although this was not done in the present work. If space permits, this may be highlighted in a revised version (but we are mindful of the recommendation of shorten the manuscript).

The above being said, we will still try and discuss the results of the model (in Section 5.1 of the original manuscript) more distinctly in terms of accumulation and ablation in the revised version.

In my opinion, the results and discussion part are not a clearly separated. In both sections, points of results and discussion can be found. Maybe you merge these two sections or make a clearer separation into results and discussion.

Thank you for this feedback. We propose to combine the Results and Discussion sections to overcome this problem and avoid repetition.

It comes far too late in the manuscript that you are using the model WaSiM. This should already be mentioned in the abstract and the introduction.

We agree, and shall alter the manuscript to state clearly in both the abstract and the latter part of the introduction section that the model was set up using WaSiM.

Snowpack simulations cannot be described as ‘actual’ measurements!

This was not our intention at all. Line 355 of the original manuscript actually stated that “*neither of the “observed” SWE time-series, which are presented in Figure 5 (alongside their simulated counterparts from the final model), are actually direct measurements.*” So we are in agreement!

Just a very general remark: Why don't you use additionally Alpine3D and compare it to your results?

Alpine3D was investigated and tested during the early stages of this project. Following that, we identified that Alpine3D does not enable the estimation of glacial dynamics or gravitational snow redistribution, factors that were believed to be important for our study and following work. It was therefore not chosen as the principal code. Whilst it would certainly have been interesting to additionally use Alpine3D and compare the snow results with those generated by WaSiM, this unfortunately lay beyond the scope of our study. As such, it represents an idea for future work, i.e. to explore in more detail whether more snow models with complexity but fewer processes (Alpine3D) or lower complexity but more processes (energy balance + redistribution in WaSiM) is to be preferred for simulations at catchment scale in steep alpine terrain.

The paper is very extensive and long. Some passages are only descriptions and repetitions of other literature. I would suggest to shorten (in total up to 3 pages) several points especially in the sections introduction, data and methods and to focus more on the relevant points of your work.

We will attempt to shorten the paper as much as possible whilst retaining information that we consider crucial. The removal of Figure 2 and Table 2, as proposed by the reviewer (see below) will also help in regard to length.

It is irrelevant which type of processing software (e.g. R) you used.

In certain instances, we feel that it can be helpful to highlight the software or packages with which certain analyses were conducted (e.g. to guide future work and to acknowledge the developers where software has been made open source). Nevertheless, in this case, we will remove all references to R (e.g. on lines 215 and 517) in the revised version of the manuscript. In any case, the software used will become apparent to readers who consult the supplementary material.

Specific:

p.6, l.147: addressed with 'd'

p.6, l.148: stations without 's'

Thank you for spotting these typographical errors. They will be corrected in the revised version.

p.6, Please insert which snow model you used. I guess you are talking about WaSiM? But can this really be described as a snow model?

The reference to WaSiM has now been added in this section (in response to an earlier point). The revised text at line 158 will read: *"Initially, a fully-distributed energy balance-based snow model that includes gravitational redistribution is established using WaSiM (Schulla, 2017) at high spatio-temporal resolution"*.

In our opinion, WaSiM is a hydrological model with a strong snow module or component. In this study, "only" the capabilities of WaSiM for interpolation meteorological data, simulating snow and ice dynamics, and estimating potential evaporation were employed. In other words, we did not proceed to simulate the remaining hydrological processes culminating in stream discharge. Essentially, since we were using these components of WaSiM as one would use a more conventional snow model, we feel this description is appropriate.

p.8, Figure 1: You mentioned the two headwater sheds Vallon de Nant and Vallon de La Vare – please mark them in Figure 1.

The extent of the Vallon de Nant sub-catchment is already marked by a polygon on Figure 1. The extent of the Vallon de La Vare is not discussed further and so actually is not really important to the comprehension of the paper. For this reason we propose simply modifying the text to indicate that the Vallon de La Vare lies immediately to the north east of the Vallon de Nant.

p.9, Section 3: In the section ‘Data’ you also describe the processing / generation of the data. Please consider renaming this section accordingly.

We will rename this section “*Data availability and processing*” to reflect the fact that this section deals with both aspects, i.e. it outlines the data that was available and/or selected for use, and also describes the processing of that data for the purposes of the study.

p.9, Section 3.1: Start by describing which input the model needs. The information is just somewhere in the following text.

Thank you for this suggestion. In the revised version, this section will start with a sentence listing the meteorological data the model needs, such as “*The model requires gridded estimates of incoming shortwave radiation, precipitation, relative humidity sunshine duration, air temperature, vapour pressure, and wind speed.*”

p.10, Eq.1: please insert a reference for Eq. 1

This equation comes from Schulla (2017). The reference will be added in the revised version (Line 229).

Section 3.2: Here information on the applied snowpack model (I guess SNOWPACK (Lehning et al. 2002) is missing. Which were the input data for the SNOWPACK simulations? Actually, you need specific meteorological input data (and not necessarily snow measurements). Why couldn’t you use the SNOWPACK simulations at the meteorological stations within your catchment? Please also think about using Alpine 3D (Lehning et al. 2006). And, if you used snow measurement data from instruments at the higher station, please specify, which snow sensors where used.

The key input data for the 1D SNOWPACK simulation, which was undertaken at the upper IMIS station (COR), consisted of the following hourly measurements:

- air temperature
- relative humidity
- wind speed
- reflected short wave radiation
- surface temperature
- snow height

A key point to emphasise is that although the model can be run using total precipitation, in which case snow height is not required, directly measuring total precipitation in such terrain is fraught with difficulty. Indeed, for this reason, the COR station only measures *liquid* precipitation directly. The measured depth therefore becomes an important input in the modelled estimates of SWE.

The list of sensors deployed at the IMIS stations is not published to our knowledge. Therefore, following receipt of this review, we requested and received the information from the SLF:

Snowheight: Campbell SR50A

Snow Surface Temperature: IR AlpuG (customized infrared sensor, http://www.alpug.ch/pdf/pdf20_IR%20Beilage2000%20d.pdf)

Snow and Ground Temperature: Campbell T107

Wind: Young 05103 (customized version for high alpine applications)

Relative Humidity: Rotronic Hydroclip
Air Temperature: Campbell T107 (thermal isolated mounting in radiation shield, not ventilated)
Reflected Short Wave Radiation: Campbell CS300
Liquid Precipitation: Campbell ARG100

Given the impact that inserting this list would have on the length of the paper (with in our view only a minor enhancement to the comprehensiveness), we propose will ask the journal before deciding whether or not to do so.

SNOWPACK could not be run at the stations within our catchment because not all of the required parameters were measured at these stations. In addition, a physically based model as SNOWPACK is very sensitive to data quality and does not tolerate gaps (e.g. in the snow depth time series). Interpolating the missing input was therefore not considered as an option.

Please see our earlier response concerning why Alpine3D results were not presented.

p.11, l.261: gradients (with ‘s’)

Thank you, this change will be made.

p.13, Figure 2: In my opinion this figure is not needed. Anyway, in Fig. 2b, the classes were not rounded to e.g. 2 digests and the classes seem to be a bit random; in Fig. 2c, the legend does not correspond to the image (no snow should be white) and to the colours in Fig. 3.

This figure was included in the original version to simply illustrate the process of developing a binary snow extent map from a Landsat image. We agree that this information is fairly standard. We therefore plan to remove this figure in the revised version of the paper. Importantly, we are happy to do this because the same information (i.e. true colour composite and delineated snow extent) will still be available to readers in Supplementary Figure 4 (for all 17 images). The length of the paper (another point raised) will also benefit from this change.

p.11, l. 282: Why did you exclude all cloud covered Landsat 8 data. Maybe there were some useful images, which were only partly cloud-covered. It should be discussed why you prefer using only cloud-free data whilst allowing a lesser spatial resolution in potential available images.

The Landsat images used were not entirely cloud free, but crucially were cloud free over our study catchments. The manuscript will be updated to highlight this (line 285). We felt that this was important because our study area is relatively small in comparison with the scale of typical cloud features. In other words, if clouds were present over some of our area, then in percentage terms this would likely be considerable.

Our scripts quantifying the match between observations and simulations and weighting scheme employed in the calibration would then have had to have been substantially altered to account for the variable image coverage. Especially given the careful procedure that was followed to define bespoke NDSI thresholds for each image, we feel that 17 images in total, spanning the full range of snow cover conditions, were sufficient to demonstrate the concept of the “granular” calibration metrics and evaluate the ability of the model to simulate varied snow cover conditions.

A final note is that excluding cloud affected images as we did of course results in a lower *temporal* (not spatial) resolution.

p.15, l.338f: IDAWEB? IMIS? Not clear to all readers please describe acronyms.

The meaning of the IDAWEB acronym does not even seem to be explained even on the MeteoSwiss website (<https://gate.meteoswiss.ch/idaweb/more.do>). Essentially, it appears to simply be a name. To

reflect this, we will change the manuscript around line 205/206, where IDAWEB is first mentioned, to read “*were downloaded from the online data portal of MeteoSwiss*”. The link to the IDAWEB page can then be found via the reference which immediately follows. Reference to IDAWEB on line 339 will similarly be changed to “the data portal of MeteoSwiss”.

In contrast, IMIS is explained by the SLF. It stands for the Swiss Inter cantonal Measurement and Information System (IMIS). This acronym will therefore be explained in the revised version (line 340).

p.15, l.345ff: Which parameters for the station was used by applying Jonas et al (2009)?

To estimate snow density according to the model of Jonas et al. (2009), the appropriate monthly parameters (*b*, *a*) for the elevation band <1,400 m were used (the station elevation being 1,146 m). Since our station lies within “Region 1” of that study, the Region 1 offset of +7.6 kg/m³ was also applied. The manuscript will be revised to make this clearer.

p.15, l.354: I strongly disagree that these are actually direct snow measurements. This has to be corrected.

Please see the earlier response (pg. 2 of this document). It was certainly not our intention to argue that these can be considered direct measurements (the contrary, in fact). We will rephrase the sentence to make this clearer. The new version will read “*neither of the “observed” SWE time-series, which are presented in Figure 5 (alongside their simulated counterparts from the final model), can be considered direct measurements.*”

Section 3.3: What is the temporal resolution of the streamflow data?

The underlying resolution of the streamflow data is rather high, being per minute. However, for the purposes of this study, the hourly mean was calculated and used. The following sentence will be inserted at line 363 stating this: “*For this work, hourly mean flows were calculated.*”

p.16, l.362: Define ‘rating curve construction’.

To address this point, Line 362 will be modified to: “an empirical stage-discharge relationship (or rating curve)...”

Also, “rating curve construction” will be changed to “the development of the rating curve”. (L366).

Section 4: As you actually describe some methods also in Section 3, I would suggest to rename this chapter and relate it especially to your modelling, calibration and uncertainty estimation ‘work’ with WaSiM.

Thank for this suggestion. Chapter 4 will be renamed to “**Numerical modelling**”.

p.16/l.365: without ‘to’

The original sentence makes sense in our opinion. But we will rephrase slight more concisely to “*...was selected as the foundation of our snow modelling approach*”.

p.19, Table 2: This table is not necessary and can be replaced by one sentence in the text.

This will be done. The revised text will read “*...each pixel was binned into one of the quadrants of a contingency matrix (a: simulated “snow” and observed “snow”, b: simulated “snow” but observed “no snow”, c: simulated “no snow” but observed “snow”, and d: simulated “no snow” and observed “no snow”)*”.

p.20, l.486ff: Is there a reference for the chosen weighting? Or why did you choose the weighting of 60:40?

There is no reference for the weighting. 60:40 was chosen so that both the spatial and temporal data had a major influence on the overall objective function, with the spatial patterns being slight dominant (the calibration based on spatial patterns being an important novelty of the study). A sentence will be added to the manuscript to make this clearer.

p.22, l.537: ‘were also developed’: I suggest to write ‘were also calculated’

We agree that the term calculated is more appropriate here, and will make then change.

**Section 5: Please rethink your subchapter captions and a general merge with Section 6.
- p.26, l.603ff: For example, this point belong more to the discussion**

Yes, as mentioned in response to a similar point raised above, we feel that it could be better to have a combined Results and Discussion section. Subchapter headings will naturally be changed to accommodate this.

- p.30, Figure 7a: Please explain more clearly what is compared and shown in this graph.

This figure simply shows the temporal relationship between the hourly simulated “snowcover outflow” and hourly observed streamflow. To enable these quantities to be equated in mm, the distributed simulations were averaged across the Vallon de Nant sub-catchment, and the streamflow measurements in m³/s were also divided by the catchment area (to give mm). The description in the text beginning at line 660 will be modified to try and make this clearer. The caption will also be lightly adjusted.

References

Cochand, M., Christe, P., Ornstein, P., & Hunkeler, D. (2019). Groundwater storage in high alpine catchments and its contribution to streamflow. *Water Resources Research*, 55(4), 2613-2630.

Warscher, M., Strasser, U., Kraller, G., Marke, T., Franz, H., & Kunstmann, H. (2013). Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps. *Water resources research*, 49(5), 2619-2637.