Dear Reviewer,

Thank you much for your review and valuable comments. Below you will find your referee comments (in black) and our responses (in blue).

With regards,

Atabek Umirbekov, on behalf of all authors

The paper addresses an important and compelling topic: the issue of choosing an adequate snow modelling scheme in the context of scarce data availability. This topic is particularly relevant for many areas of the world where instrumentation and monitoring is rather poor, yet the population depends on meltwater resources. The authors presented a machine learning-based model that requires simple and/or commonly available input data and no calibration. The model showed good performances in reproducing SWE both in the subset of stations not used for calibration and in two other remote, orographically complex and scarcely monitored stations. The model structure, training, validation and limitations are well explained and clear. The validation is extensive and considers point-wise and large-scale cases.

My suggestion is a major review. The motivations are the following. Generally, throughout the paper, I often found the literature review either insufficient or even absent. The description of the data used is scattered throughout the text, which doesn't help clarity. Figures often lack axes ticks, labels and/or units.

We are grateful for your comprehensive feedback on the manuscript and the valuable critical comments you provided. In response, we will enhance the literature review and expand the discussion of the important aspects that you have highlighted in your comments both here and below. We agree with your observation that the current version of the manuscript presents a mixing of data and methods, and we are committed to reorganizing them for clarity. Additionally, we will redesign incomplete figures and improve their overall organization, as you've suggested in your comments.

The comments are the following:

--- *MANUSCRIPT* ---

0. General comments:

0.1 I suggest adding a comprehensive “Data” section where the authors can (a) list all the data they used, separating them in subsections for model training and validation, point-wise and large-scale; (b) roughly describe the geography/orography/data availability for the datasets they chose.

As requested, we will gather information on data used for both model training and evaluation under a separate sub-section “Data”, and provide brief details on climate and topographical characteristics. The repositories indicated in the Data availability section will be updated with data and script used for large-scale evaluation of the model.

0.2 I suggest restructuring the final part of the paper with a freestanding “Model limitations” section and a “Conclusions” section encompassing and enhancing what is now in section “Summary”.

As requested, we will separate ‘Model limitations’ into standalone section, and add “Conclusions” section to the manuscript

0.3 I suggest a re-reading and improvement of the English language, there are syntax/grammar errors in the text and the structure of some sentences is confusing (see comments for each section). Please check that the used tense is consistent along a section or paragraph.

0.4 Notations: throughout the text, figures and tables, please make the Celsius degree symbol consistent (°C); correct the Elevation unit from m to m a.s.l.; when a quantity is non-dimensional (i.e. NSE), please use the non-dimensional unit ([-]).

We will edit and improve clarity of the text across those highlighted sentences, and bring unit notations into consistency

1. Introduction

I suggest rewriting the Introduction by significantly expanding the state of the art and literary research, taking into account the following comments:

- L30: Suggested citation: Beniston M. (2008), Extreme climatic events and their impacts: Examples from the swiss alps. In: Díaz HFRJ (ed) Murnane, climate extremes and society. Cambridge University Press. New York. USA. pp: 147-164.

  Thank you for suggesting an appropriate reference for this sentence. We will refer to Beniston, 2008 in this line

- L31-39: This paragraph generally lacks references and examples on both kind of models; I suggest providing a small literature review.

  Thank you for this suggestion. We will provide a more detailed review of both types of the models and add supporting references (such as Essery, 2020; Jonas et al., 2019) .

- L37: "… research often opt for relatively simpler conceptual TI models…" references and examples are needed.

  We suggest to refer to Hock, 2003 and Ohmura, 2001 for this sentence.

- L40-41: I find this sentence too general and poorly supported by literature (the authors only provide one example). For example, in this recent study https://doi.org/10.5194/hess-26-3447-2022 the authors showed how a PB snow-hydrological model substantially outperformed a conceptual TI model. Both models were applied on the same spatial domain (catchment Dischma), and the TI model completely missed the snowmelt-induced discharge timing (see Figure 7 d-e).

  Thank you for pointing at the issue of insufficient references. We suggest to amend the sentence and supplement it with the following references: *"both types of models can produce similar results when calibrated for the current climate and applied to the same spatial domains (Bavera et al., 2014; Magnusson et al., 2011; Shakoor et al., 2018)."*
  In addition, we will add a new sentence into the paragraph: *"Models calibrated to the same conditions in the current climate can produce different predictions under climate change (Carletti et al., 2022)."*

- L51-60: I find this paragraph dedicated to the state of the art preceding the authors' work too short and general. I suggest expanding this section by better detailing the findings of previous works (upon which the authors rely for their work) and the critical issues of the previous works (which the authors seek to address in this paper).

  We will expand overview of machine learning applications for snow modelling as requested.

2. Model description

- The default threshold temperature value for rain/snow separation is set to -1 °C. Here, it would be necessary to justify this choice, or at least provide references, because this tuning parameter can vary a lot in snow/hydrological modelling (see for example https://doi.org/10.3390/cli9010008 for a TI model and https://doi.org/10.5194/hess-26-1063-2022 for a PB model).

  Thank you for this suggestion. We will add additional description with regard to Ts threshold, such as the following:

  "The Ts constraint differs from classical temperature-based partitioning methods where the threshold defines precipitation in a binary way, as either 100% rainfall or 100% snow. In this model, snow-precipitation partitioning is simulated using its inherent learned relationships, but only until the temperature drops below Ts. At that point, any precipitation is regarded as 100% snow. For example, when the average temperature (TAVG) is 0°C, depending on other climatic and topographic variables, the model is likely to simulate a significant portion of precipitation as snow (around 75% of precipitation), even if Ts is set at -1°C."

- L82-85: *"… and is available as a set of functions […] respectively"* If the subject is "a set of functions", then verbs should be "calculate" and "generate". Otherwise, the sentence as it is is unclear and I suggest rephrasing, dividing or better explaining.

  Indeed. We will correct the sentence accordingly.

- L110: *"As it was noted above, the SVR model has two tunable parameters: cost and gamma…"* Actually, gamma is never mentioned. The authors mention "sigma" on L99. Please clarify.

  We apologize for this confusion. We meant the same parameter, 'gamma', which is sometimes referred in literature as 'sigma'. We will stick to term 'gamma' throughout the text

3. Model validation

- L160: Please cite https://doi.org/10.1016/0022-1694(70)90255-6

  Thank you for suggesting the reference. We will make a reference to Nash and Sutcliffe 1970 in line 160

- L180: As mentioned in Comment 0.1, Mendoza and Western Pamir are not mentioned earlier in the text as data used for validation and are only introduced here.

  Mendoza Andes and Western Pamir will be introduced in a new section 'Data'

- L199-200: Do the authors refer to Figure 4? If so, Figure 4 needs to be mentioned. See the comments about Figures.

We appreciate this suggestion. We will ensure that all figures are appropriately introduced and referenced in the mansucript.

- L202: *"… the rain-to-snow transition modelled using the metadata of the 520 validation SNOTEL stations."* Do the authors mean that there are observations/data on the transition between rain and snow for all the 520 stations? And how was that used in modelling? Please clarify.

  The main motivation behind this analysis is to have an understanding how the model simulates precipitation-snow partitioning during snow accumulation phase. The following new exert will provide additional details in this regard: *"Since the SNOTEL observations do not contain explicit information on precipitation-snow transition, we decided to use a sample of the dataset to simulate the transition depending on climate inputs (temperature variables) and topographical characteristics (e.g. Elevation). More specifically we have filtered the SNOTEL observations that closely fall on this phase by selecting observations that meet the following non-exhaustive main criteria: 1) observations for October or November when precipitation is non-zero 2) average temperature (TAVG) is less than 10 or higher than -10°C, 3) accumulated SWE is less than 20mm. We then run the model using the obtained sample of observations and estimated solid fraction of precipitation simulated by the model, i.e. amount of dSWE estimated by the model in respect to precipitation amount."*

- L206: *"… does not exceed 100%"* do the authors mean does not *reach* 100%?

  Yes, indeed, 'not reach 100%' is more appropriate and we will rephrase this part accordingly. Thank you for this correction.

- L210: I suggest justifying this sentence with a plot or a better explanation. Again, if this information is contained within some metadata, this needs to be explicitly stated.

  The histogram on the bottom left of Figure 7 could serve as supporting plot. We will make an appropriate cross-reference in the sentence.

- L241: How did the authors calibrate Ts? Please clarify.

  We calibrated Ts for each of the stations with the objective of maximizing the Nash-Sutcliffe Efficiency of the model`s simulations with respect to observed SWE. We will include this clarification into the text.

- L255-256: Can the authors verify this assumption? Shortly after, in the text, the authors write the same for the SnowMIP station SNB, so I assume it is possible?

We could use the example of the calibrated Ts values for the stations located in Sierra-Nevada as supporting evidence for this assumption; although some of the stations in this domain are located in close vicinity to each other, the calibrated Ts values exhibit high variability ranging from -1°C to -5°C). The main message of the sentence is that arbitrary altering Ts may lead to overcalibration through the error compensation effect. We will point to this issue in the Model limitations section.

- L292: The authors should explain the meaning of *"class balance accuracy"*.

  We will supplement this sentence with an explanation of class balance accuracy.

4. Model sensitivity and uncertainty assessment

- L305: Is there a reference for this method? If so, I suggest adding it.

  Yes, this method is explained in Fisher et al., 2018 and Greenwell et al., 2018. We will supplement these references into the sentence.

- L311: *"... depending on the phase considered ..."* Do the authors mean "precipitation phase"? Please clarify. Also, the reference is missing.

  We refer to two general phases of snow metamorphosis: snow accumulation and snow ablation. We will make this clearer in the revised version of the manuscript. Our apologies for the missing reference; it was supposed to be a cross-reference to the Figure 10 further down.

- L316: What do the authors mean by *"relative comparison"*? Please clarify.

  In the given context, "relative comparison" means that the importance of those topographic variables is made in relation to other variables used by the model. We will rewrite this line in the text to make it clearer.

- L349: Please refer to Table 1 when addressing the different model settings.

  A cross-reference to the Table 1 will be included in the L348-349.

- L355: What do the authors mean by *"when outliers are controlled for"*? Please clarify.

  The boxplots in Figure 12 show extreme limits, which exclude outliers. More specifically, the minimum and maximum limits of the boxplots are determined by (1st Quartile - 1.5 * IQR) and (3rd Quartile + 1.5 * IQR), where IQR represents the interquartile range (Hu, 2020). To prevent confusion, we suggest that we remove the phrase *'when outliers are controlled for'* from the sentence.

5. Summary

- L375: The concept of equifinality is only addressed at the end of the paper but it is never mentioned earlier. The most important papers on equifinality are not cited (see https://doi.org/10.1016/0022-1694(89)90101-7, https://doi.org/10.1016/0309-1708(93)90028-E, https://doi.org/10.1016/j.jhydrol.2005.07.007). If overcoming equifinality is one of the aims of the paper, this needs to be addressed in the Introduction and also in the discussion of the results. And additionally, how does the model improve equifinality? This needs to be explained and justified. The results shown in Figure 12, for example, seem contradictory to this sentence, because there the authors show that one can obtain similarly good model performances with different sets of parameters.

  Thank you for suggested references. We will introduce issue of equifinality in snow modelling in the introduction and expand its discussion in respective part of the manuscript.

  In this sentence we refer to the challenge of calibrating multiple parameters in hydrological and snow modelling. This challenge is particularly prominent in hydrological modelling, where even relatively simple snow modules require calibration of at least two parameters: the precipitation-snow threshold and the degree factor. Considering that there are many other parameters for different components of a hydrological model, it would be easy to end up with multiple combinations of optimal model parameters. On other hand while our model contains only one tuneable parameter (Ts), it shows generally plausible performance in diverse climatic and topographic conditions upon using the default value of Ts. We hypothesize that replacing the snow module with a model that is based on generalizable empirical relationships may help to reduce the equifinality issue, especially when employing conceptual hydrological modelling.

  Figure 12 shows performance of four GEMS models that differ in a number of required inputs, but contain only a single parameter (Ts) which can be adjusted. All four models' performances depicted in figure 12 were obtained by using the default value of the Ts (-1°C)

  L383-385: This sentence is not clear. What do the authors mean by *"instrumental"*?

  We will edit the sentence, by replacing '*instrumental*' with '*helpful*' or '*useful*'. Here we meant that "*balance (in) complexity, data requirement, and transferability… could be helpful for operational monitoring and hydrological modelling in data scarce domains*."

- L385: Similarly for the equifinality, the problem of finding empirical relations and parametrizations is never addressed before in the text. If this is one of the aims of the paper, it needs to be addressed in the Introduction accompanied by proper references (as parametrizations of different kinds are already widely used in snow/hydrological modelling).

Thank you for raising this. We now recognize that the statement in this sentence may have been too assertive and requires further verification. We will remove this sentence from the manuscript.

- Please consider mentioning the undercatch selection issue within the Model limitation section.

  By filtering observations for precipitation undercatch, we assume that the evaluation dataset is comparatively free of this issue. However our selection algorithm also filtered records where inconsistencies between accumulated precipitation and SWE may be reasoned by wind-induced snow-drift. Disentangling these two phenomena is challenging without further research. The model cannot capture/simulate snow-drifts, we acknowledge this limitation in lines 263-265 and explicitly stated it in lines 407-408.

--- *FIGURES* ---

General comments:

- When a figure is composed by different subplots, as it is often the case in this paper, something that enhances clarity very much is naming each subplot differently, for example with letters like (a), (b)… And then, throughout the text, referring to each subplot like Figure 5a, Figure 5b etc.
- I suggest improving the figure referencing generally and throughout the whole text: often the authors describe the results referring to specific subplots of a same Figure by only mentioning the general Figure once at the beginning of the paragraph. Referring to each specific subplot before introducing each finding highlighted by the subplot increases clarity significantly.

  Thank you for these recommendations. We will review the organization of the figures accordingly, and ensure they are properly introduced and referenced in the text.

Specific comments:

- Figure 2: Axes ticks and labels (latitude, longitude) are missing, legend is missing.
- Figure 3: Axes labels are missing.
- Figure 6: Left plots: missing adimensional symbol for NSE ([-]), missing unit for snow meltout date error (days?), missing y-axis label. Right plots: Missing axes ticks and labels (latitude, longitude).
- Figure 7: Same as above.
- Figure 8: y-axis label and units are missing.

- Figure 11: “Latitude” is spelled wrong, missing units, missing y-axis ticks and labels.

  Thank you for pointing out at these deficiencies. We will correct and align the figures accordingly.

References:

Bavera, D., Bavay, M., Jonas, T., Lehning, M., & De Michele, C. (2014). A comparison between two statistical and a physically-based model in snow water equivalent mapping. *Advances in Water Resources*, *63*, 167–178. https://doi.org/https://doi.org/10.1016/j.advwatres.2013.11.011

Beniston, M. (2008). Extreme climatic events and their impacts: examples from the Swiss Alps. In H. F. Diaz & R. Murnane (Eds.), *Climate Extremes and Society* (pp. 147–164). Cambridge University Press. https://api.semanticscholar.org/CorpusID:129970616

Carletti, F., Michel, A., Casale, F., Burri, A., Bocchiola, D., Bavay, M., & Lehning, M. (2022). A comparison of hydrological models with different level of complexity in Alpine regions in the context of climate change. *Hydrology and Earth System Sciences*, *26*(13), 3447–3475. https://doi.org/10.5194/hess-26-3447-2022

Essery, R. (2020). *Understanding and getting started with physically based snowmelt models*. The International Association of Hydrological Sciences. https://iahs.info/uploads/Commissions/ICSIH/ICSIH Understanding physically based snowmelt models.pdf

Fisher, A., Rudin, C., & Dominici, F. (2018). *Model Class Reliance: Variable Importance Measures for any Machine Learning Model Class, from the “Rashomon” Perspective*. https://api.semanticscholar.org/CorpusID:126316805

Greenwell, B. M., Boehmke, B. C., & McCarthy, A. J. (2018). *A Simple and Effective Model-Based Variable Importance Measure*. 1–27. http://arxiv.org/abs/1805.04755

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, *282*(1), 104–115. https://doi.org/https://doi.org/10.1016/S0022-1694(03)00257-9

Hu, K. (2020). Become Competent within One Day in Generating Boxplots and Violin Plots for a Novice without Prior R Experience. *Methods and Protocols*, *3*(4). https://doi.org/10.3390/mps3040064

Jonas, T., Mcphee, J., Skiles, M., & Marks, D. (2019). *Understanding strengths and limitations of temperature-index snowmelt models* (Issue October). The International Association of Hydrological Sciences. https://iahs.info/uploads/Commissions/ICSIH/ICSIH snow modeling article FINAL.pdf

Magnusson, J., Farinotti, D., Jonas, T., & Bavay, M. (2011). Quantitative evaluation of different hydrological modelling approaches in a partly glacierized Swiss watershed.

Hydrological Processes, *25*(13), 2071–2084. https://doi.org/https://doi.org/10.1002/hyp.7958

Ohmura, A. (2001). Physical Basis for the Temperature-Based Melt-Index Method. *Journal of Applied Meteorology*, *40*(4), 753–761. https://doi.org/https://doi.org/10.1175/1520-0450(2001)040<0753:PBFTTB>2.0.CO;2

Shakoor, A., Burri, A., Bavay, M., Ejaz, N., Ghumman, A. R., Comola, F., & Lehning, M. (2018). Hydrological response of two high altitude Swiss catchments to energy balance and temperature index melt schemes. *Polar Science*, *17*, 1–12. https://doi.org/https://doi.org/10.1016/j.polar.2018.06.007