In black: Referee comments
In blue: Authors' response

RC1: 'Comment on egusphere-2025-1214', Anonymous Referee #1, 24 Mar 2025

The temperature index method is a convenient and widely used method in snow simulation. This study estimates the important parameters in the temperature index method based on the published SWE and climate datasets, and analyzes the influence factors of these parameters. The results can provide insights into the temperature index method, making this paper worth publishing in HESS. Having said that, I would like to point out some major concerns that should be addressed before publication.

Dear referee, we thank you for your fast, thorough and constructive review of our paper. We appreciate you finding the paper worth publishing in HESS, and **we will gladly address the concerns** you have pointed out to make the study better. See below detailed responses to your comments.

1. Potential Circular Logic: The temperature threshold and melt factors are estimated based on the SWE dataset, which is subsequently used to evaluate the performance of the temperature index model. This approach risks circular reasoning. A more rigorous method would involve dividing the dataset into two subsets—one for parameter estimation and another for model validation.

We agree that some circular logic could be interpreted here, as the SWE dataset is used both for parameter estimation and for evaluation. As Sections 3.3 and 3.4 describe, there are three different parameter sets and model simulations for each station in the study (illustrated in Figure 2 and Table 2 in the manuscript). We respond to your point separately for each parameter set.

The first set simulates SWE with the same parameter values for all stations across the Northern Hemisphere. The aim of this is to evaluate how one single parameter set performs when applied to all stations across the Northern Hemisphere. It is true that the melt factor value used (3.64 mm/°C/day) is the median of all empirically derived melt factors. We have randomly subsampled ½ of the dataset 1000 times and the computed median melt factors ranged from 3.61 to 3.66. Therefore, we doubt the results would change if we did that for this first set of simulations.

The aim of **the second set** of model simulations is to evaluate the model performance using empirically derived parameter values for each station individually. In other words, we evaluate the best possible parameter set, based on observations. Here, splitting the dataset spatially is not possible as a parameter needs to be estimated for each station. Instead, **we will split each station temporally** and use half the data for empirically deriving the parameter, and the other half for model performance evaluation.

In **the third set** of model simulations, we use estimated parameter values from multilinear regression models based on climate variables. Here, we agree that we should spatially split the dataset in two subsets. The first subset (⅔ of the available stations) will be used to build the regression model that estimates model parameters based on climate variables, and the second subset (⅓ of the available stations) will be used for model evaluation. **We will modify this and include the new results in the revised manuscript.**

2. Subjectivity in Determining the Melt Threshold: In Section 4.1.2, the melt threshold appears to be assumed as 0°C, with supporting analyses provided. However, if the threshold were slightly adjusted around 0°C, the conclusions in Section 4.1.2 would still hold. To minimize subjectivity, a quantitative approach should be employed. While the melting process is expected to occur at 0°C from a physical standpoint, the temperature data used do not precisely reflect the conditions at the exact location where phase changes occur (e.g., the snow surface for melting and the atmosphere for precipitation partitioning).

We agree that there is subjectivity in determining the melt threshold. In the literature, this threshold varies between -1°C and 1°C, so we heuristically chose 0°C as the central value of this range, and we supported it with our analyses in Section 4.1.2 (Figure 5 in the manuscript). However, in this section a quantitative approach is not plausible due to data limitations. This limitation is discussed in lines 161-166 (Section 3.3), and owes to the fact that the NH-SWE time series cannot confidently distinguish if a SWE decrease (at the daily time scale) is due to snowmelt, snow redistribution, sublimation, or a data error, and therefore limits the capacity for this more quantitative approach.

Nevertheless, the point of the reviewer remains extremely valid, and our analyses indeed support that a mean daily temperature of ~1.1°C might be more appropriate for accurately estimating the onset of the snowmelt season (see Figure B2), as this may better reflect the snowpack becoming isothermal. On the other hand, a varying melt threshold temperature will add a free parameter to our approach and add complexity to our broader aim to provide robust estimates for melt factors across the Northern Hemisphere. **We will analyse the sensitivity of model performance to a varying melt threshold temperature and we will include it and discuss it in the revised manuscript.**

3. Clarity in Time Scale: The study computes temperature thresholds and melt factors at the daily scale in some instances, while averaging them in others. This inconsistency makes the methodology difficult to follow. A clearer distinction between different time scales should be provided.

You are right that the time scales across sections and figures might be confusing. The analyses of the two melt thresholds and the melt factor start at the daily time scale by comparing the daily time series of temperature and precipitation with the daily time series of SWE. This leads to the insights on temperature thresholds and melt factor at the daily time scale from Figure 3, Figure 5, and Figure 6a. However, in this study we use temperature-index modelling with constant parameters (non-varying in time, only in space), therefore we compute the median of the daily temperature thresholds and melt factor to obtain a seasonal value for each year (e.g. Figure 6b). The calculated seasonal values are further averaged to obtain a mean temperature threshold and melt factor for each station (that is Figure 4, Figure 6c, Figure 7, Figure 8). In the revised manuscript, **we will make this clearer in the methods section, as well as in each subsection and each Figure.**

4. Effectiveness of a Single Parameter Set: The model employing a common single parameter set outperforms the other two models in certain aspects. This raises the question of whether complex parameter estimation methods are necessary. A discussion on the added value of these methods compared to a simpler approach would be beneficial.

We agree that a more thorough discussion about this is needed in the manuscript, and we find it encouraging that a simple parameter set has good performance and provides a useful justification for studies with no or little physical information on which to base their model. On the other hand, we also want to understand and provide some physical basis for the spatial variability of the melt factors and their performance. There are a variety of reasons leading to small differences in model performance between the three parameters sets. First, the melt temperature threshold is the same for all sets of simulations, which limits the possible differences in results. Second, the snowfall temperature threshold varies per station but only minimally, as we set a minimum value of 0°C (for which we have now added a justification, in response to one of your minor issues from L257). For the empirically derived and estimated parameter sets, 79% and 86% of stations have a threshold value of 0°C, respectively. For the rest of the stations, values are mostly below 2°C (see Figure 1 below). The small variability of the snowfall temperature threshold across parameter sets may also contribute to the small differences in model performance across parameter sets. Note that the figure excludes stations with a snowfall threshold of 0°C.

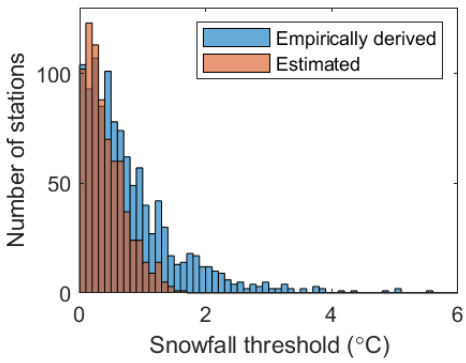

Figure 1. Distribution of the snowfall threshold in the empirically derived and estimated parameter set, excluding stations with a 0°C threshold (which represent 79% and 86% of stations for the empirically derived and estimated parameter set, respectively).

The last factor of model performance variability across parameter sets is the melt factor. This is the parameter that varies most between stations and between parameter sets, as seen in Figure 2 left panel below. The melt factor differs between the two parameter sets (Figure 2 right panel) because our parameter estimation model based on climate variables has low predictive skill (L275 in manuscript).

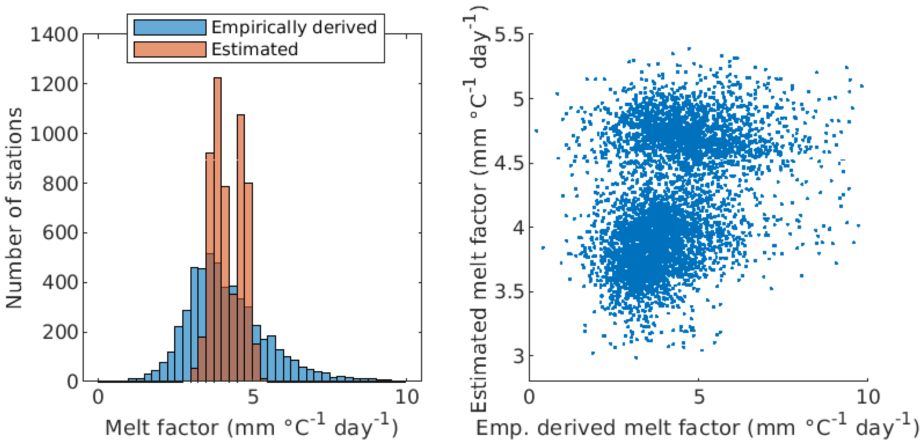

Figure 2. Melt factor distribution in the empirically derived parameter set vs the estimated parameter set based on climate variables.

The model performance variables most sensitive to the melt factor are the melt rate and the time of the end of the snow season. However, there is a strong correlation in model errors between the two sets of model simulations (see Figure 3 below). This indicates that model performance is not very sensitive to model parameters, at least when evaluated over long-term time series, as we do in this study.

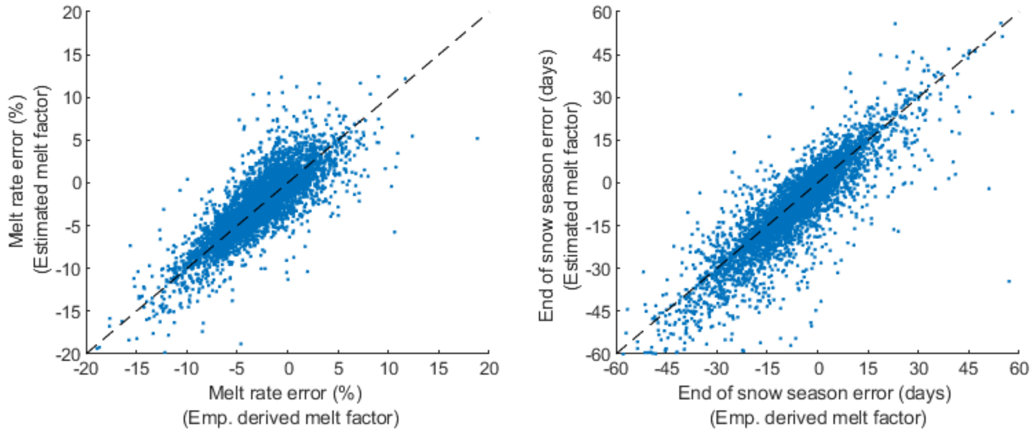

Figure 3. Model errors in the empirically derived parameter set vs the estimate parameter set based on climate variables. Melt rate error (left) and timing of end snow season error (right). Dashed line indicates the 1:1 line.

As you point out, this suggests that for long-term studies using temperature-index modelling, it is not necessary to use complex parameter estimation methods, as the use of commonly applied values in the literature will lead to similar model performances. **We will add this discussion to Section 5.3.**

Minor issues:

Thank you for pointing out these minor issues. We will incorporate your suggestions accordingly.

L106: Provide the full name of SNOTEL. Will do.

L123: Ensure consistency in terminology (e.g., CHCN-d, CHCNd, and CHCN-Daily). Will do.

3.1: Make it clear what you are specifically referring to here. There are more than two indices in the Table 1. We will specify whether each term is a snow season term or a climate index.

L162: “toe” should be “to”. Indeed.

L163-166: Difficult to understand. Please rephrase and explain it more clearly. Will do.

3.4: Consider merge sections 3.2 and 3.4, moving the descriptions of the model to the beginning of 3.4 section.The rationale behind this decision was that it was confusing to start describing the estimation of model parameters, without having initially described the model we are using.

Figure 3d: what does the dark orange mean? - This is just the overlap of orange and blue bars. We will add this detail in the caption.

L257: Why is a threshold lower than 0℃ not allowed? - We assume the reviewer means L235? Due to our definition of snowfall threshold in lines 152-156 in the manuscript, our snowfall threshold estimation model predicts threshold temperatures below 0℃. However, based on Figure 3 in the manuscript, precipitation below 0℃ is very likely to be snow. We therefore only allow the snowfall threshold to be 0℃ or higher, to avoid capturing too much snowfall as rain.

In some figures, the text is overlapped. Please check and modify them. Thank you for pointing this out, we will revise the figures.