Responses to Reviewer #1

This manuscript presents how a simple off-line snow temperature index model (B-TIM) can be considered to highlight discrepancies between snow water equivalent (SWE) products from three reanalysis (JRA-55, ERA5 and MERRA-2) and an additional product (ERA5Snow) for historical period (1980-2020). The authors used either biased, or adjusted temperature and precipitation from reanalysis as input data for B-TIM. The SWEs produced with B-TIM and various sets of input data were then compared to the SWEs produced by the reanalysis. Climatological characteristics and interannual variability were investigated. To carry out this study, they improved and translated the previous version of B-TIM and made it publicly available. This manuscript opens up the possibility of using a simple off-line model for large-scale snow cover studies.

We thank the reviewer for all the comments and have included responses along with requested additional calculations in the Author Comments. Line references included here refer to the original manuscript.

Some modifications in the structure could lighten the text and focus more on the results and the contribution of using a simple off-line model like B-TIM. For example, a large part of section 2.1 would have a more appropriate place in the SI. After all, this is not a paper about improving B-TIM, but more about using it. If this is not the case, please change the title and specify this aspect more clearly in the objectives.

The primary objective of the model description section is to thoroughly document the settings of the B-TIM before analyzing its output and using it. When model decisions are not defined explicitly, it can be challenging for others to reproduce those decisions and/or compare across datasets fairly. To this end, we propose to move Tables 1 and 2 to the SI while maintaining the written text describing the time evolution of the modeled snow. This way, the section will be shortened without breaking up the model description.

It would be useful to remind the reader of the context in the results. For example, simply add a sentence to remind us that ERA5 and ERA5Snow have the same meteorology, which explains why we don't have BrE5S.

Thank you for the suggestion to clarify this point. The following text will be added after L270: Two of the reanalysis datasets, ERA5 and ERA5Snow, share the same meteorology. Therefore, there is only one BrE5 dataset that is produced.

Methodological choices (e.g., bias adjustment) could be justified in greater detail, with more references where possible.

The methodological choices stem from the hypothesis that major differences in SWE across reanalyses are consequences of the mean bias in the temperature and precipitation, as is currently described in L308-309. This motivation could be strengthened in the bias adjustment section.

We propose an addition around L309: If biases in mean meteorological conditions are the primary source of snow bias, a correction toward similar climatological conditions should yield more similar modeled snow. We implement a first-order correction which relies on monthly correction values derived from climatological temperature and precipitation conditions (see SI).

In the literature, correction factors similar to ours are typically derived from in-situ observations for the purpose of model calibration, validation, or sensitivity testing (e.g. Cho et al., 2022 or Raleigh et al., 2015). Our application is different in the sense that we are adjusting the forcing data in both directions between pairs of datasets to investigate whether we can bridge the gap between them and then attribute the differences to the mean biases. Some of this additional context will be added to the discussion.

The results focus on very large domains (Northern Hemisphere, Eurasia, North America). It would have been interesting to look at these results more regionally.

Regional studies are of major interest, especially with respect to a breakdown according to land type. The purpose of this paper is to characterize snow over simplified terrain and at large scales, but the two accompanying papers Mudryk et al. (in discussion) and Mortimer et al. (in discussion) provide some regional results and perhaps should be more clearly highlighted here.

We propose to use the following to replace the text starting in L77: This study has as its focus hemispheric snow. Even at these large scales and excluding the complicated case of mountain snow, there are discrepancies that should be characterized. For exploration into regional performance, there are two other studies prepared for publication, Mudryk et al. (in discussion) and Mortimer et al. (in discussion), which include all the datasets discussed in this paper. Mudryk et al. (in discussion) compares a suite of 23 snow datasets, ranks the performance of each one, and examines the inter-dataset consistency. Mortimer et al. (in discussion) presents an expanded SWE dataset that combines in-situ and airborne SWE measurements and assesses snow dataset performance when compared to the in-situ record. This in-situ data is used across all three studies. Our main scientific work…

Temperature and precipitation bias were corrected with a multiplicative factor. As precipitation is a zero-bound variable, it is generally corrected by a multiplicative method, whereas temperature is often corrected by an additive method. The choice of this method is not sufficiently justified. Can you explain thoroughly why you chose a multiplicative factor and why you apply it this way?

The methodology requires some more explanation, which we outline here and will endeavour to integrate into the paper. Commonly, when correction factors are based on the bias between ground-truth and modeled data, normally distributed biases are corrected with an additive method, and lognormally distributed biases are corrected by multiplicative adjustment. This usually means temperature biases are corrected by adding a constant, while precipitation biases are scaled multiplicatively. This has the added benefit of maintaining the zero-bound of precipitation. The application here is different. We use the multiplicative factor for both variables because:

- Climatological *temperature* biases between datasets are small (no more than 2 % of the absolute temperature) year-round. For fractionally small differences compared to the raw value, a multiplicative method and an additive method yield very similar results.*
- We did some filtering to ensure that no erroneously large or small corrections were applied (e.g. from division by a small precipitation value)
    - Precipitation scaling factors were limited between 1/3 and 3, and wherever dry conditions prevailed in either the native or target dataset (daily average precip. <0.2 mm), a scaling factor of 1 was used (no adjustment).
    - Temperature scaling factors were limited between 0.99 and 1.01
- Using the same method simplified the experiment runs

*Explicit comparison of the two methods:

Given Dataset 1 (D1) and target Dataset 2 (D2), both of which are functions of location and time, let climatological temperature conditions for a given month be represented by C1 and C2. With a multiplicative bias correction as in the paper, the temperature at time step *t* is adjusted as in (1).

$$D1_{mult}(t) = D1(t)\frac{C2}{C1} = D1(t) + D1(t)\left(\frac{C2}{C1} - 1\right) = D1(t) + D1(t)\frac{(C2-C1)}{C1} \qquad (1)$$

Meanwhile, an additive method corrects D1 following (2).

$$D1_{add}(t) = D1(t) + (C2 - C1) = D1(t) + D1(t)\frac{(C2-C1)}{D1(t)} \qquad (2)$$

The correction terms differ because D1(t) and C1 are not identical. However, both methods yield the right climatology, $\overline{D1_{add}} = \overline{D1_{mult}} = C2$.

If we write D1 in terms of departures from the climatology C1, we can rewrite the scaling factor in equation (2). The two scaling factors are close as long as the sub-daily variations $\delta D1$ are small compared to C1.

$$\frac{(C2-C1)}{D1(t)} = \frac{C2-C1}{C1+\delta D1(t)} = \frac{C2-C1}{C1\left(1+\frac{\delta D1(t)}{C1}\right)} \approx \frac{C2-C1}{C1}\left(1 - \frac{\delta D1(t)}{C1}\right) \qquad (3)$$

Though the scaling factors are similar, multiplication can stretch or compress the diurnal cycle -- this can theoretically affect snow accumulation when temperatures are near freezing (primarily near the margins) but it should be a reasonably small effect. Given the reasoning above, we do not expect a large difference between the correction methods, and a preliminary test has been run to support this.

Figure captions contain results, whereas captions should only contain descriptions of the elements present in the figure (colors, symbols, etc.). Please remove the result part in the captions.

The captions of Figs. 2, 3, 6, 7, 8, 9 have been edited to remove results. They were all already described in the text.

L102: Did you perform tests regarding the 20% of precipitation reduction?

The 20% precipitation reduction is a value based on estimates of snow loss that were incorporated and tested during model development (Brown, 2003 and personal communication). The original validation was based on optimizing agreement with in-situ data at locations in eastern Canada. Given some recent increases in the availability and quality of in situ SWE, snow depth, and snow density information, future B-TIM development may revisit this 20% reduction. However, it is outside the scope of our current study and we did not test it.

Example citation for new validation data: Vionnet et al., 2021 or Mortimer et al. (in discussion).

L109: Table 2. Please find a more consistent way to present column "Model variable". For example: « t2m (ID 167) » instead of « Parameter ID 167: "t2m" ». Add the model variable name for SWE in ERA5Snow. Also, this table could go in the SI, as it doesn't provide much relevant information to the text.

Thank you for the suggestions. The intention was to copy the descriptions as directly as possible from the modeling centers, but a standardized approach is likely clearer. We will update Table 2 and move it to the SI, with reference to it in L102.

Table 2, L215, L268, L428 : Modify MERRA2 to MERRA-2.

Thank you for identifying these, they have all been fixed.

L232: To take advantage of the fact that you have an SI, it might be interesting to present the differences in domains used for the different reanalyses (land grid points, mountainous grid points, etc.).

It is an interesting suggestion to look at different domains. We hope the new pointers to the evaluation papers Mudryk et al., (in discussion) and Mortimer et al., (in discussion) will be sufficient for interested readers to turn to for regional analysis. It is best to interpret the strengths and limitations of these products in a larger ensemble.

Fig. 3 : Please describe colors used in the scatterplots in the caption; modify ERA5-Snow to ERA5Snow; consider using hatches for ERA5Snow in the right panel and present the legend in a neutral color.

We have improved the clarity of Fig. 3 with consistent labeling and both colour and hatching to distinguish ERA5Snow and ERA5.

Please note the following change: the values presented in this figure are slightly different from the previous version (no changes to discussion/conclusions). This resulted from mistakenly using Nov-May data to produce the figure in the original manuscript. This version correctly uses Nov-March data.

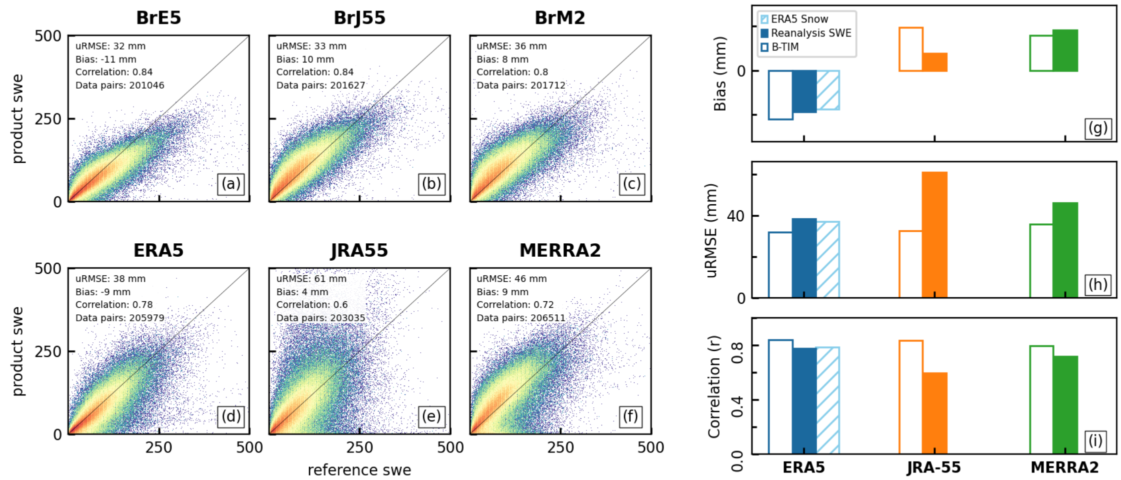

Fig. 3 SWE product validation against snow course and gamma SWE measurements. Figs. 3a-f consist of scatterplots showing all valid data pairs (snow course, product) from November to March over 1980-2018. Summary statistics, including the bias, unbiased root mean squared error (uRMSE), and correlation, are included in the legend and are summarized in Figs. 3g-i. uRMSE is calculated by removing the mean from the reference SWE and each set of product SWE values, then calculating the RMSE with those datasets.

In L293 we will include the following description: The scatterplots (figs. a-f) are coloured as heatmaps to display the concentration of points, which is highest where the colour is red. Each scatterplot represents over 200 000 pairs of points.

L374: Modify JRA55 to JRA-55.

Fixed.

L469 : Modify ERA5-Snow to ERA5Snow.

Fixed.