Dear authors,

Thank you for revising the manuscript and for your responses. Since one point of critique in the first round (open discussion phase) was serious ("insufficient model validation") and the revised MS has changed quite substantially, I decided to ask a third, new referee for an independent opinion. This person has a strong background in regional atmospheric modeling. You will find his/her review attached, and you will see that the referee has major concerns with the study and rated the scientific quality of the paper as "poor".

This referee points out a "lack of realism", which links directly to the aforementioned "insufficient model validation", suggesting that the paper has not improved convincingly in that respect. In particular, the uncertainty in precipitation is emphasized by the new referee, which also connects to Dieter Scherer's comment "...the entire study depends on the accuracy of downscaled precipitation. It would therefore be of utmost interest to better understand the uncertainties in the WRF output." from the first round of review. In addition, my own review (see attached) raises a question in the same direction (are model and observations in agreement? See point **(2)** of my comments). Therefore, revisions are necessary, and I will make a final decision after the re-submission and your responses whether the manuscript can be considered for publication or not.

I hope these comments are helpful for clarifying your study further.

Thomas Mölg
Handling Editor &
Co-Editor-In-Chief TC

Editor Review of revised MS

My own comments on the revised version concern two areas.

(1) Readability
Below I list ~20 examples that I caught during my reading, which illustrate a lack of precision in statements or procedures. While each case on its own is probably a minor problem, all together make reading the paper quite hard. Please go through the paper carefully and revise these and similar problems to enhance the unambiguity of statements. I hope you can see that the below examples are not reader-friendly.

96-100: Several instances of "temperature"; it is unclear whether they refer to air or the glacier (surface).

101-105: "snow" and "snowfall" are used here; do they refer to the same (I guess you mean snowfall as you refer to solid precipitation)?

123: "amount of water"; all phases or just a particular one?

181: RMSE of 1.8 °C; were annual, seasonal, monthly values used in the calculation? It is not obvious.

Figure 2 caption: What is the "nearest WRF grid" in this context? Do you mean "nearest grid point"?

Figure 2: The figure is supposed to show a station/WRF comparison. How do panels (b) and (e) fit here? They show a station trend.

198: RMSE of 11.4 mm/month; were annual, seasonal, monthly values used in the calculation? Again, it is not obvious (per month can also be used as unit for mean seasonal or annual values).

Figure 2: it is referred to a lot in the text, but almost always without specification of which panel is meant (a few times I understand that the entire figure is addressed, but that can't be the case always).

232: is "figure 3 in de Kok et al., 2018" meant? Figure 3 in the present paper is a temperature figure.

Figure 3: Time step for the correlation calculation not clearly specified (annual, season, …).

Figure 4: Time step for the correlation calculation not clearly specified (annual, season, …).

279-281: How can changes (increase/decrease) be inferred from Figure 8, which shows the mean diurnal cycles?

Figure 7: The captions says regions with "growing" and "shrinking" glaciers, but the legend shows WKSK and SW HMA. Are these two definitions exactly the same regions?

Figure 9: Is the scale bar adjusted to the min/max values in the maps? I can't see much dark blue in the maps.

332: "snow" or "snowfall"; I guess you mean the latter?

382-385: Same as above, which time step is used for calculating the correlations? "Interannual" can also compare seasons or months between years (or annual values).

Discussion starting in 409: "roughly matches" is not clear enough. Also, your Figure 6b shows snowfall trends, yet you refer to precipitation here. The cited reference also shows precipitation. Please be consistent for comparisons.

General: What tests are used to determine the p values?

General: Are the trends tested for significance?

General: with the addition of new data sets, mixing up Methods and Results has become more serious than in the first version. The readability would benefit from having more descriptions of data and technical procedures before the results section (Section 3).
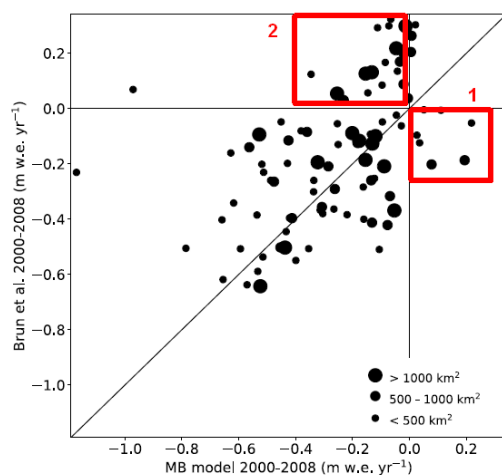
(2) Scientific Contents

My main topical comment refers to Section 3.3, where one key message is that areas of growing or shrinking glaciers are consistent in model and observations. While this would be a nice result, I assume that readers will have trouble understanding it when they look at Figure 10. In particular, where the model region tends to be positive (marked Box 1 below) the Brun values are mostly negative, and where the model region tends to be negative (Box 2) the Brun results tend to be positive. One could also conclude that model and observation show the opposite with regard to neutral/stable mass balances. This discrepancy adds to the referee assessment of "a lack of realism".

Minor

87: represent 95% → needs a reference

158: "almost complete" is redundant

170: extracted or downloaded instead of "collected"

249: "our model" → which one is meant? The trajectory model?