Dear Reviewer,

Thank you for your thorough evaluation of our manuscript. We take good note of your main comments about the geographical sensitivity of NARR model forcings and the need to offer more insights into the transferability of the approach to other glaciers. Saskatchewan glacier is indeed favorably placed close to a NARR grid node, which may have contributed to the success of the downscaling and ensuing mass balance modelling. We will follow some of your recommendations to address this point, i.e., by calculating downscaling statistics for the other surrounding NARR gridpoints as well as a spatially interpolated NARR to the AWS location using bilinear interpolation. This should provide interesting and more transferable insights. A discussion of the differences between the coarse NARR topography and the AWS location and its effect on NARR biases was already presented in the discussion, but it will be further developed and clarified after the new sensitivity tests are completed.

As for the downscaling of wind speed: the current model structure does not allow spatially variable wind speed, which is a limitation. However and as you noted, the NARR wind speed did not represent well the wind speed at the AWS, due to the katabatic wind regime not being capture by the NARR product. This was discussed in the MS, but we indeed did not quantify the impact of this uncertainty on the mass balance model. We thus propose a novel approach to test the model sensitivity to using NARR instead of AWS observations, by forcing the mass balance model with AWS observations for the observation period 2014-2016 but replacing, one at a time, each driving variable by its downscaled NARR counterpart. This should highlight how uncorrected NARR errors (i.e., errors other than the bias corrected by downscaling), impact the simulated mass balance.

Regarding the downscaling nomenclature: we have used bias-correction methods (scaling, also compared to EQM) to downscale the coarse NARR variables to the location of the on-glacier AWS. These downscaled NARR variables were then either distributed to the glacier grid using lapse rates (air temperature and precipitation) or assumed constant (RH and wind speed). These are fairly standard practices, even if probably simplistic for wind speed. The sensitivity tests proposed above should at least address the sensitivity to NARR errors for each driving variable. I believe you use the term `downscaling` to refer to the extrapolation of point data (from the AWS location) to the glacier grid. We used the...
term ‘distributing to the glacier surface/grid’ as we think this step is separate from downscaling NARR to the AWS location, but we will revise closely the nomenclature to clarify the methods as per your comments.

The text will be restructured after including the new results, addressing your suggestions for each subsection.