Dear Editor,

We thank the reviewers and editor for considering our revised manuscript and for the positive response to the changes. Please find attached the final version of the manuscript where technical corrections have been addressed.

In the following response, reviewer comments are indicated in black and our responses are indicated in blue italic font. We have included excerpts of the changes made in the submitted manuscript. No other changes have been made.

We would again like to express our sincere appreciation to the reviewers and the editor for their time and efforts in reviewing our manuscript.

Sincerely,

Kamilla H. Sjursen and co-authors

Reviewer 1

Following the first round of reviews, the authors improved the readability of several sections of text, clarified their reasoning, moderated strong claims, added a feature importance analysis (appendix C) and refined the description of their data cleaning and selection approach (appendix A, B). Several other suggestions were addressed in the responses to the reviewers.

The feature importance analysis is a welcome addition to the manuscript. I accept the work as is but would like to raise one point for consideration in future research. In L670 you note that correlation can cause some features to appear less important when assessed with permutation importance. However, in the case of SVF, permuting this feature actually decreases the MSE across all months, which implies that removing it could improve model performance. This apparent contradiction — improved performance when SVF is permuted (Fig. C2) despite its relatively high weight (Fig. C1) — may suggest overfitting. Because XGBoost does not automatically exclude weak predictors, all variables remain available for splitting. A feature with little or no true relationship to the target can still appear frequently in the trees, particularly if the training setup is not strongly regularized. This can artificially inflate its weight. It may therefore be informative to compare MBM performance on the training and test sets to assess whether overfitting is happening, and whether additional feature optimization could mitigate it. While in this case study the effect seems minimal, it could become more pronounced in other regions or applications.

All in all, I deem this version of the manuscript fit for publication.

We thank the reviewer for assessing our proposed changes and for the positive response to these. We would again like to express our gratitude for the thorough and constructive review of the manuscript, which we believe has greatly improved its quality.

We would also like to thank the reviewer for the valuable insights on feature importance, and agree that this is an important point that should be mentioned. In the submitted manuscript, we have included a comment on this in the feature importance analysis in Appendix C.

> We caution against placing too much emphasis on the specific details of the feature importance analysis. For example, when assessing permutation importance, correlated features (i.e. skyview factor and slope) may appear to be less important since, even if one feature is permuted, the model can rely on a second correlated feature. However, the high weight of the skyview factor (Fig. C1a), in combination with a slight decrease in performance for many months with permutation of this feature (e.g., Fig. C2a–e), may indicate some overfitting. Although the effect seems to be minimal in this case, it highlights the need for careful feature optimization when using correlated features. Overall, the main findings of the feature importance analysis presented here are consistent across metrics and physically meaningful with respect to the main meteorological drivers of mass balance on Norwegian glaciers.

Below some final minor textual corrections that may be considered:
- L 43: reword “alleviated the lack of” to “alleviated the shortage of”

Done.

> The increasing availability of geodetic mass balance observations has recently alleviated the ~~lack~~ shortage of glacier-specific observations. These observations assess glacier surface elevation changes from time series of satellite-derived digital elevation models (DEMs) over decadal time scales (e.g. Dussaillant et al., 2019; Shean et al., 2020; Hugonnet et al., 2021). Most

- L59: There is some repetition here with the addition of the new text. I suggest removing “including to unsurveyed glaciers and years” from L55 as this is now explained in more detail in the following sentences.

Done.

> from training data and apply them to make accurate inferences on new, independent data. They can thus learn statistical relationships between mass balance components and topographical and meteorological variables that are transferable across space and time ~~, including to unsurveyed glaciers and years~~ (e.g. Guidicelli et al., 2023). This means that ML models can leverage sparse in situ data, such as annual and seasonal glaciological measurements, to provide high spatio-temporal resolution

- L302: I suggest moving “respectively” earlier in the sentence, as it is not clear whether you are referring to point/elevation-band mass balance vs. mass balance gradients, or seasonal vs. annual.

Done.

> We compare predictions from all models (MBM, GloGEM, OGGM and PyGEM) to available glaciological and geodetic mass balance observations for glaciers in the test dataset over the common modelling period 1980–2019. In Sect. 5.2.1 and Sect. 5.2.2, we respectively consider point/elevation-band mass balance and mass balance gradients, ~~respectively,~~ on seasonal and annual time scales. Glacier-wide mass balances are compared in Sect. 5.2.3 on monthly to decadal time scales. We evaluate

- L406: The addition of the feature importance discussion and the more cautious statement regarding MBM downscaling are great. Minor comment: the wording “The key to this ability” is somewhat confusing and could be rephrased.

Done.

> of precipitation and elevation difference features in winter months (Fig. C2a–c and k–l), suggests that it is able to downscale precipitation locally. The same is true for temperature in the summer months (Fig. C2e–i). The key to this ~~ability~~ downscaling capacity lies in using the elevation difference between the stake and the climate model as a feature (Fig. 3)~~which enables~~, enabling MBM to effectively map the relationship between climate and elevation.