Authors' Response to Reviews of

A minimal machine learning glacier mass balance model

Marijn van der Meer, Harry Zekollari, Matthias Huss, Jordi Bolibar, Kamilla Hauknes Sjursen, Daniel Farinotti.
The Cryosphere,

RC: ***Reviewers' Comment***, AR: Authors' Response, ☐ Manuscript Text

Dear Editor,

We thank you for your time and thoughtful evaluation of our manuscript, "A minimal machine learning glacier mass balance model" [Paper EGUSPHERE-2024-2378]. We greatly appreciate the positive feedback from you and the reviewers. In response to your technical comments, we have revised the manuscript accordingly, with changes clearly highlighted. Below, we address your comments that require further explanation. All minor editorial suggestions have been directly incorporated into the revised version.

1. Editor

1.1. Technical correct #20

RC: ***L349: "identify" instead of "validate"?***

AR: L349 read

> Our results illustrate that this approach enhances performance by identifying the predictors that most effectively explain PMB observations in Switzerland. Consequently, these findings not only demonstrate the utility of this ML method but also help validate the climatic drivers commonly associated with glacier MB [citations to studies].

By "validate," we meant that our study's results confirm existing empirical knowledge about the drivers of MB, but this time using a different approach—namely, machine learning. To make this clearer, we changed the sentence to the following:

> Our results illustrate that this approach enhances performance by identifying the predictors that most effectively explain PMB observations in Switzerland. Consequently, these findings not only demonstrate the utility of this ML method but also help validate the climatic drivers of glacier MB commonly identified by glaciological studies [citations to studies].

1.2. Technical comment #1 & #2

RC: ***Caption of Fig. 3 and L115-116 and: Captions usually do not discuss the figure content, but rather briefly describe the sub-panels with specific references within the main text. For instance, L1-8 of Fig. 3 caption should be included in the main text (e.g., L115-116) with specific references to sub-panels a, b and c where appropriate. The figure caption can be brief: "Conceptual overview of … (ML) model. a) Pre-processing … from MeteoSwsiss. b) Example of an X array with a half-yearly aggregation, i.e., including four predictors (n=4), and c) the resulting predicted PMB timeseries (^y) covering N years."***

RC: ***Caption of Fig. 4: Similar comment to the one above.***

AR: We know that the captions for Fig. 3 and 4 are particularly long, but this is because these figures illustrate the most complex processes in the study. We wanted to ensure that readers could fully understand the figures by reading the captions without needing to refer back to the main text. But thanks to your comment, we realized that the references to Fig. 3 might not be optimally placed. To improve clarity, we decided it would be more effective to move the "Experiments with the model's input" section, which discusses the various forms of input aggregation, ahead of the architecture and training of the model. This adjustment ensures that by the time Fig. 3 is first referenced, readers will already have all the necessary context to understand it.

References to Fig. 3 now come first as follows:

> *3.1 Experiments with the model's input:* [...]
>
> To reduce the predictor space of miniML-MB, we rely on temporal aggregations of monthly climatic data. Here, the input features $\vec{X}$ of miniML-MB for a site are given as an array of dimension $N \times n$ (Fig.3b), where N is the number of annual PMB observations, and n is the number of predictors made from aggregates of temperature and precipitation (ranging between 2 and 24). These aggregates are computed from monthly MeteoSwiss measurements using the mean for temperature and the sum for precipitation. We explore four levels of temporal aggregation (Fig.3a):

We also removed redundant information from the caption of Fig. 3:

> Conceptual overview of the training of miniML-MB, the point surface mass balance (PMB) machine-learning (ML) model. For each PMB measurement site i, miniML-MB is trained to simulate PMB from meteorological variables. (a) Pre-processing of monthly air temperature (T in $^\circ$C) and total precipitation (P in $\mathrm{m\,w\,e}$) from MeteoSwiss. (b) Meteorological variables are formed into $\vec{X}$, an array of N rows (number of annual PMB measurements at the site), and n columns (number of predictors). ~~Predictors are temporal aggregations of monthly variables at different resolutions: annual (2 predictors, $n = 2$), half-yearly ($n = 4$), seasonal ($n = 8$), monthly (no aggregation, $n = 24$), or optimal seasonal ($n = 2$). During optimal seasonal aggregation, each column aggregates consecutive months (e.g., mean T of Oct.–Jan. and total P of Feb.–Apr.).~~ This example shows $\vec{X}$ during half-yearly aggregation with four predictors ($n = 4$, summer and winter half-years). (c) $\vec{X}$ is given as input to miniML-MB which predicts PMB $\vec{y}$ covering N years.

And similarly for Fig.4, now its subpanels are referenced in the text,

> *3.3 Training and testing:* [...]
>
> During this framework, in independent testing, the dataset is shuffled and split into five folds (subsets) (Fig.4a). Each fold is used once as an independent "test set", unseen by miniML-MB during training, while the model is trained (or "fitted") on a "training set", which is the remaining aggregate of folds (hyperparameter tuning, see below). This process is repeated five times until miniML-MB has made predictions for each "test set", and these are aggregated to recreate a time series covering each year for which PMB measurements were taken (Fig.4b).

and the caption of Fig.4 captions was shortened:

Testing framework of miniML-MB illustrated at site P2 on the Plattalva glacier. (a) At each point surface mass balance (PMB) measurement site i, miniML-MB makes PMB predictions, which are then evaluated using a cross-testing framework. Input climate predictors and observed PMB measurements are divided into five subsets. Five times, miniML-MB is trained on four of these subsets and makes predictions on the remaining (unseen) test subset. (b) The predictions made on these five test subsets (one colored dot per subset) are aggregated to reconstruct a PMB time series for the site, covering all years with observed data. The accuracy of these predictions is assessed against the observed PMB (gray lines).~~ using metrics such as mean absolute error, root-mean-squared error, and Pearson correlation. This figure illustrates this evaluation process at site P2 on the Plattalva glacier~~