

Response letter

Spatio-temporal reconstruction of winter glacier mass balance in the Alps, Scandinavia, Central Asia and Western Canada (1981-2019) using climate reanalyses and machine learning

Guidicelli, M., Huss, M., Gabella, M., and Salzmann, N.

Dear Editor,

We sincerely thank the editor and referees for taking the time to review our manuscript and providing constructive feedback to improve our manuscript.

In response to the detailed comments by the two referees we have thoroughly and substantially revised our manuscript. Below, please find a brief summary of the most important changes of the study and related changes in the manuscript:

- We carefully revised the terminology used in the manuscript (snow accumulation, SWE, winter mass balance), which is now more coherent with the methods that have been applied. Consequently, this has substantially improved the flow of the manuscript.
- We improved the clearness of the goals (winter mass balance reconstruction) and methods (derivation of vertical profiles of the winter mass balance). Substantial changes have been made to the text in the abstract, introduction and method sections.
- We discussed the role of the year as predictor in the gradient-boosting regressors (GBRs), and the related implications.
- We described the meaning of the hyperparameters defining the GBR.
- We further discussed the choice of applying a GBR instead of other machine-learning methods.

We are confident that these changes have significantly improved our work so that it finally will reach the needed quality and standard to be accepted and published in 'The Cryosphere'.

On behalf of all co-authors,

Matteo Guidicelli

Comments by Anonymous Referee #2

Legend: Referee's comments; *authors replies*

Manuscript excerpts with [added text](#) and ~~deleted text~~

I thank the authors for the thorough revision of their manuscript, they have addressed the concerns that I raised. I see 2 minor additional complements, that would likely further improve the manuscript without requiring much effort.

We would like to thank the referee for the positive feedback and for the additional minor complements that have helped us to further improve the manuscript.

1/ the use of the downscaled reanalysis variables as predictors should be better explained in the Methods (section 3.1 for instance) and also on Fig 3. The methods state that "In addition to the original variables, the GBR requires some downscaled variables of the reanalyses as predictors at the glacier elevation, ..." but when reading the legend of Table B3 it is less clear whether for instance all single level variables from Tables B1 and B2, are used in addition to the ones mentioned in Table B3, or just some of them. Following, we do not know which of the reanalysis variables (downscaled? all?) are withdrawn in the tests from Fig 3.

All variables listed in Table B3 are required by the GBR. This is now indicated at the beginning of the Methods section (lines 147-148):

[The list of predictors required by the GBR is summarized in Tab. B3, while the variable names are described in Tabs. B1 \(ERA-5\) and B2 \(MERRA-2\).](#)

and at the beginning of Section 3.1:

In addition to the original variables; [\(all "constants", all "single level" and all "land surface" variables in Tabs. B1 and B2\)](#), the GBR requires some downscaled variables of the reanalyses as predictors at the glacier elevation [intervals \(see Tab. B3\)](#), including air temperature, dew point temperature and relative humidity (for MERRA-2 and ERA-5), vertical velocity of air motion (for ERA-5 only) and specific humidity (for MERRA-2 only).

Furthermore, the caption and the content of Table B3 have been modified in order to make clear that all "single-level" variables and "constants" of Tables B1 and B2 are required by the GBR. Here, we report the new caption of the table:

[Summary of the variables used by the GBR models: all constants and single-level variables from the original reanalyses \(Tabs. B1 and B2\), variables downscaled at the elevation of the \$B_w\$ data \("variable_down"\), mean of the variables during the accumulation season considering only days with a minimum precipitation of 5 mm \("variable_Ppos"\), difference between "variable_down" and the variable at the original grid of the reanalysis \("delta_variable"\), year, glacier coordinates, slope and aspect. \$P_{grid}\$ is the precipitation of the original reanalysis. The variable names have the same roots as those reported in Tabs. B1 and B2.](#)

The groups of predictors withdrawn in the tests of Fig. 3 (new Fig. 8) have been described in the caption, which is now consistent with Table B3. Here, we report the new caption of Fig. 3 (new Fig. 8):

Overall root mean squared error (RMSE) between B_w and GBR models using different groups of predictors, for all analyzed glaciers and years: (a) ERA-5 site-independent GBR, (b) MERRA-2 site-independent GBR, (c) ERA-5 season-independent GBR, and (d) MERRA-2 season-independent GBR.

~~*Topography The "pressure levels" group refers to all variables derived from the pressure-levels data, i.e. "downscaled" variables and "delta variable" (except for the elevation difference) in Tab. B3. The "single level" group refers to all "single-level" variables listed in Tab. B3. The "topography" group refers to the topographical parameters describing the reanalysis's subgrid complexity (all "constants" in Tabs. B1 and B2) and the average slope and aspect of the glaciers by using the information provided in the Randolph Glacier Inventory version 6 (RGI Consortium, 2017). For the list of variables included in each group of predictors see Tables B1 (original ERA-5 variables), B2 (original MERRA-2 variables) and B3 (downscaled ERA-5 and MERRA-2 variables).~~

2/ Fig 3 and its analysis : it is at first very surprising that the suppression of reanalysis data, leads to the best performance/smallest rmse for the season independant GBR, as reanalysis variables could be hypothesized to carry the "climatological" information shaping the temporal dynamics of the adjustment factors.

The authors analyse this finding in the new section "5.2.1 Site-independent and season-independent GBRs", whereby seasonal-independant GBR models are hypothesized to split into individual submodels adapted to individual glaciers and hence better able to predict a glacier-specific (but maybe temporally constant ?) adjustment factor inferred from the training accumulation seasons. Actually, I see a possible alternative explanation of this finding, which occurs concomitant with the important role of the year as a predictor for season-indep GBR : it may point towards the fact that the year may be the one predictor to actually vehicle the climatological information (the GBR model then learning 'by heart' which years correspond to negatively biased winter accumulations for the glaciers in this region, and which other years correspond to more balanced/less biased ones).

Though such a model may perform well in the past, its use in extrapolation to future conditions (using for instance GCM/RCM data in lieu of reanalysis), is doomed to fail.

Maybe the authors could comment on this and add a note in the discussion if they find this explanation possibly relevant.

Many thanks for the interesting interpretation of the role of the year as predictor for the GBR model. This is certainly a possible explanation of its importance and was only briefly mentioned in Sec. 4.2 (a short sentence has been added):

However, skipping reanalysis variables from the set of predictors leads to higher errors for some individual glaciers, especially in the representation of the temporal variability of the ~~B_w data snow accumulation~~. In fact, excluding the reanalysis variables, the year is the only predictor able to vehiculate the climatological information, in other words, the year is the only predictor that could be used to predict a different adjustment factor depending on the accumulation season (all the other predictors are constant in time).

We discussed the impact of the use of the year as predictor in Sec. 5.2.1 and Sec. 5.2.2.

In Sec. 5.2.1 (Site-independent and season-independent GBRs), we extended the interpretation of the results of Fig. 3 (new Fig. 8) by including your comment, which we think that completes our previous explanation (lines 425-435):

In general, the year was included in the set of predictors for the GBRs because it may allow learning the climatological information and potential trends in terms of reanalysis biases against the B_w data. This can be relevant in case reanalysis variables included in the predictors are not able to represent the hypothetic trends of these biases, which can exist due to the increasing availability of observational data that could have been assimilated by reanalysis models over the years. From Fig. 8, we notice that the year has a different impact on the site-independent and the season-independent GBRs. For the site-independent GBR, the withdrawal of the year from the predictors leads to a smaller increase of the overall RMSE than the withdrawal of the reanalysis variables (pressure levels and single level). In contrast, for the season-independent GBR, the best performance is reached by withdrawing the reanalysis variables. Thus, given the mentioned ability of the season-independent GBR to potentially create sub-models for specific glaciers using their coordinates, we think that the

year may be the one predictor to actually vehiculate the climatological information in this case (the GBR model learns which periods correspond to larger/smaller biases for the tested glacier).

In Sec. 5.2.2 (Spatial and temporal transferability of the GBR models), we discussed the potential use of the season-independent GBR in the extrapolation to future conditions, related to the use of the year as predictor (lines 461-466).

The inclusion of the year in the set of predictors is not problematic for temporal reconstruction of B_w with limited gaps, but could be a limitation for extrapolation to future conditions (using for instance GCM/RCM data in lieu of reanalysis). In fact, even though the use of the year as predictor could allow identifying and modelling trends of reanalysis biases against the B_w data, these trends would be limited to the training period of the GBRs. In contrast, the exclusive use of reanalysis variables as predictors could allow identifying and modelling trends of biases as function of specific climatological conditions represented by reanalysis variables.

Comments by Referee #3: Maussion, Fabien

Legend: Referee's comments; *authors replies*

Manuscript excerpts with *added text* and ~~deleted text~~

Dear authors,

I would like to start my review with an apology for the time it took me to write it. I am aware that endless review processes are a real strain (especially for PhD students), and some unplanned personal matters prevented me from writing sooner.

Another (less important) reason for the late review, however, is that the paper is a difficult read. I really enjoyed the study and I find that it is relevant and interesting, but as of the current version the manuscript feels like a puzzle that the readers have to solve by themselves, because a lot of relevant information is scattered across the manuscript. I also believe that the first round of reviews shaped the manuscript in a way which makes it less readable now. I think however that the manuscript can be brought in a reasonable shape with some restructuring.

Many thanks for your detailed review, which helped us to again sharpen the focus of our study, and to make the entire study more linear – and thus easier to read –, and more robust and consistent with our goal.

I will start with the most important point: what is the purpose of the paper? The title says “Snow accumulation over glaciers ... inferred from climate reanalyses and machine learning”. But, in reality, snow accumulation is never analyzed (or even plotted! I only see correction factors everywhere). What is analyzed is the capacity of a statistical model to reconstruct winter mass-balance (not snow) in space and time.

The term snow accumulation or SWE were used under the assumption that snow melt is generally negligible for the considered period and high elevations. However, we agree that it is more correct to refer to winter mass balance than snow accumulation, especially when we discuss the output of our models. Now, this has been modified in the title and coherently throughout the entire paper (incl. abstract and introduction) to better guide the reader.

Note that the study is well introduced: the problems stated in the introduction are real, and it IS a good idea to use winter mass-balance observations to look at biases in reanalysis data. At the end of the introduction, however, the authors state: “we thus aim at providing improved observation-independent SWE estimates at highest elevations of different mountain ranges across the Earth”. But the manuscript does not provide anything like that, does it? I saw no SWE data, and I also didn't see a code & data availability section (against TC's policies, by the way: https://www.the-cryosphere.net/policies/data_policy.html).

I think that this is the main problem of the manuscript, as the reader is left wondering what the paper is about. Some themes which (I feel) are developed throughout the manuscript:

- Training a statistical model to reconstruct winter mass balance (WMB) from partial information
- What information is needed to do so successfully, and what problems are occurring when data becomes scarce (this is, in my view, the most interesting aspect of the study)
- What are probable bias in winter precipitation in reanalyses
- What are the differences between MERRA-2 and ERA5 (although, to be honest, I don't recall much discussion of this point despite the fact that having two datasets significantly clutters many figures in the manuscript).
- The WMB elevation profiles and how your model can sometimes reproduce those (to my surprise)

What is not developed in the manuscript:

- The difference between different statistical model choice (this is a bit of a weakness as it make the paper very descriptive, but is not a big issue in my opinion)
- Whether or not the method developed in this study will be used to develop regional products. This is very important because if yes, the paper needs to be a bit more careful in its wording as suggested by Reviewer #2. The paper is already much better at discussing limitations, but I think that if the plan is to derive actual products from the method, the abstract needs to state that this is the goal and to be more precise about what's needed to reach this future goal.
- If the goal is not to make some sort of product in a future paper, then I would like to suggest going back to my point above about clear study motivation statements.

This may sound like harsh comments, but I do not intend them to be that way: I think that the study has potential! It would be very beneficial to the paper to be more clearly written, to better explain what is done and why. I'll do my best to provide a more timely review at the next iteration.

Thank you very much for your clear indications and remarks. Our replies regarding the introduction and the goals of the paper are provided by following your further specific comments below.

We also added the data availability section to the paper. Some contents have been moved from the acknowledgments section to the data availability section, which has been included at lines 600-608:

Data availability.

Winter mass balance data separated per elevation intervals are freely available at https://wgms.ch/data_databaseversions/ (EE-MASS-BALANCE data sheet in WGMS, 2021, last access: 01.06.2021). The average slope and aspect of the glaciers were obtained from the Randolph Glacier Inventory version 6 at <https://www.glims.org/RGI/> (last access: 05.05.2021, RGI Consortium, 2017). ERA-5 hourly data on single levels and on pressure levels were downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store at <https://doi.org/10.24381/cds.adbb2d47> (last access: 01.06.2021) and <https://doi.org/10.24381/cds.bd0915c6> (last access: 01.06.2021), respectively (Hersbach et al., 2018b, a). MERRA-2 Land Surface Diagnostics, MERRA-2 Single-Level Diagnostics and MERRA-2 Analyzed Meteorological Fields data are available at <https://doi.org/10.5067/RKPHT8KC1Y1T> (last access: 13.06.2021), <https://doi.org/10.5067/VJAFPLI1CSIV> (last access: 13.06.2021) and <https://doi.org/10.5067/A7S6XP56VZWS> (last access: 13.06.2021), respectively (Global Modeling and Assimilation Office (GMAO), 2015b, c, a).

Specific comments

- I still don't think that the title reflects the content of the paper well (see general comments)

The title has been changed in order to better reflect the content of the paper. The period reported in the title and in the manuscript has also been modified for coherence with the most recent winter mass balance data available from the WGMS dataset at the moment our study was conducted. In fact, the previous period (1981-2021) was related to analyses performed in the original version of the manuscript (climatological trends) but the related section has been removed after the first round of reviews.

~~Snow accumulation over glaciers in the Alps, Scandinavia, Central Asia and Western Canada (1981-2021) inferred from climate reanalyses and machine learning~~

Spatio-temporal reconstruction of winter glacier mass balance in the Alps, Scandinavia, Central Asia and Western Canada (1981-2019) using climate reanalyses and machine learning

- Introduction: clearly state the objectives of the study, and what will be shown in the paper. Why are these regions / glaciers chosen, etc.

Substantial changes have been made to the introduction and the abstract in order to better highlight the motivation and the contents of the study, as well as to justify the selection of the regions/glaciers analyzed in the study. We also make clear that the method developed in the study will not be used to develop regional products (in response to the general comment above), indicating that this is an exploratory study (now, this is also indicated in the abstract (line 6)). Here, we report the new last paragraph of the introduction section:

In this study, we thus aim at ~~providing improved observation-independent SWE estimates~~ analyzing total precipitation biases of reanalysis datasets (ERA-5 and MERRA-2) over the snow accumulation season on glaciers, i.e. at the highest elevations of different mountain ranges across the Earth. ~~The precipitation estimates are compared with the winter glacier mass balance data covering a period of up to 39 years from 95 glaciers in the European Alps, Scandinavia, Central Asia and Western Canada. The selection of these regions/glaciers depends on the data consistency and availability (see Sec. 2.2). Ultimately, we aim at reconstructing the winter glacier mass balance from partial information. In order to achieve this goal, we develop and evaluate a machine-learning approach based on gradient boosting regressor (GBR) models (see Friedman, 2001) to adjust the total precipitation of reanalysis dataset (ERA-5 and MERRA-2) over the accumulation season. The GBR model is then used to derive SWE estimates over glaciers in the Alps, Scandinavia, Central Asia and Western Canada. Data on snow accumulation distribution at the end of the accumulation season covering a period of up to 41 years from 95 glaciers were used to train our approach (main driver of snow accumulation) along the elevation profiles of the glaciers.~~ More specifically, the GBR models aim at allowing the spatiotemporal transferability of the learned information over the 95 glaciers to other glaciers with no ground observations and/or filling gaps of observational series. The new information provided by our exploratory study is expected to be helpful to improve the calibration of glaciological and hydrological models in observation-scarce regions.

- Line 144: the motivation and implication of using total seasonal averages needs to be discussed in depth. Intuitively, a model using temporal information (even at the monthly scale) would perform better, but I understand that this is not feasible in this context.

The motivation is given by the absence of reference data at higher temporal scales (e.g. monthly) and at high elevations. The data compilation provided by the WGMS based on dozens of detailed national monitoring programs is unique in terms of providing consistent in situ data in different regions of the world at very-high elevations, i.e. on glaciers, that can be used for comparison with precipitation datasets. This comment has been added at the beginning of Sec. 2.2 (winter mass balance data, lines 106-108). We think that jointly with the reformulation of the goals of the study (where Sec. 2.2 has been cited), this is now much clearer:

The data compilation provided by the WGMS based on dozens of detailed national monitoring programs is unique in terms of providing consistent in-situ data in different regions of the world at the highest elevation of mountain ranges, i.e. on glaciers, that can be used for comparison with precipitation datasets.

The implication of using total seasonal averages has been discussed at the end of Sec. 5.2.2. Given the context of the study, we could only speculate that the performance would be better by using information at higher temporal scales, however, such information is not available and we could not support our comments with any data. Thus, we rather focused on the description of the models' limitations (lines 467-474):

Furthermore, the use of total seasonal averages as inputs for the GBRs has limitations for the application of the presented approach. The GBRs provide an estimate of the overall adjustment factor for the reanalysis precipitation according to the average climatic conditions of the accumulation season. Thus, a unique adjustment factor is estimated for the whole accumulation season. The application of this adjustment factor to daily precipitation data, would allow obtaining an average estimate of the SWE evolution during the accumulation season. However, the obtained SWE evolution would neglect potential melt of snow during the accumulation season (see Sec. 2.2), as well as

potentially different adjustment factors for precipitation at higher temporal scales than seasonal. This limitation is related to the lack of reference in-situ observations at higher temporal resolution than seasonal available for our study, focusing in different regions of the world and at high elevations.

- The methods section feels incomplete. I truly don't understand how your model is actually able to simulate WMB profiles, because my understanding at the end of the methods section is that you use seasonal totals of climate predictors to simulate total seasonal WMB of glaciers. It's not clear to what purpose "downscaling" is used, and to what elevation the variables are downscaled (I assumed the average glacier elevation). Are you using elevation bands to reconstruct WMB as a function of elevation and then average per area somehow to get the glacier specific WMB? Where is this procedure described? Or, do you actually use elevation band data from WGMS for training? You can see that I'm confused.

For each glacier we used the WGMS winter mass balance data separated per elevation intervals. All elevation intervals have been used in the training of the GBR models, which are thus able to provide estimates at different elevations for the same glacier. This information has been better highlighted in different parts of the manuscript:

- *Abstract and Introduction: we specified that reanalysis precipitation is adjusted along the elevation profiles of the glaciers (lines 10; 76-77)*
- *Data (Sec. 2.2): we explain that we used the winter mass balance data separated per elevation intervals (lines 109; 112)*
- *Methods (first paragraph, Sec. 3.1, Sec. 3.2): we repeated both indications multiple times (lines 145; 157; 160-161; 167-168; 188)*

We also added further information at the end of Sec. 3.2.2 (lines 229:231):

During the training phase of our models, the B_w data were weighted by considering the area of the glacier related to the respective elevation interval. Larger glaciers (and elevation intervals related to larger areas) thus receive more weight in the training of our models than smaller glaciers (and elevation intervals related to smaller areas).

However, the majority of the results shown in the study are based on glacier-wide averages as indicated in Sec. 3.3.2, where we show how the glacier-wide averages are computed.

- L175: to be honest I don't think the benchmark is very fair, because the parameter K seems to be a parameter to tune for each reanalysis / situation. It is also not data informed at all. I am not requesting to change this at this stage, but I personally don't put much value in this benchmark.

Thank you for this comment. We agree that the comparison between the benchmark and the GBRs is not very fair (for the season-independent GBR especially). We included the benchmark in order to show what we can expect by applying a simpler and completely not data-informed model in the case of the selected glaciers. We think that this can help the reader to understand the added value of the GBRs in terms of overall bias and vertical profiles for winter glacier mass balance. The benchmark shows that for glaciers with a non-continuously increasing B_w along the profile, a more sophisticated model is needed. In order to avoid potential misinterpretations, we specified more clearly that the benchmark method is not data-informed (at the beginning of Sec. 3.2, line 168).

This has been also remarked in the Discussion (Sec. 5.1.1, lines 378; 390).

- L204: "For these cases, groups of data in the 10-fold cross-validation contain data of different years but different groups can contain data of different years of the same glacier." -> this is really unclear. I assume only one glacier is used each time? Are you therefore building 95 models (one for each glacier) here? After reading the rest of the manuscript I see its not, but I really wonder what value there is to interpolate in time with a model that is trained on highly inhomogeneous data, and it seems that the data with the most explanatory power is obviously the data on this very glacier.

We have rewritten this paragraph in order to make clearer how both site-independent and season-independent GBRs are trained and tested. Here, we report the new paragraph for readability (lines 207:217):

The validation and the test data were defined differently depending on the goal of the GBR model. For both reanalysis products (ERA-5 and MERRA-2), we built two different GBR models with two different goals and two different cross-validation and test schemes.

The first GBR model is site-independent and aims at "extrapolating" the B_w data in time and space (over glaciers with no B_w data). For the site-independent GBR, we built 95 models, one for each glacier, trained and validated with the data of the other 94 glaciers. Each glacier is used, in turn, as an independent test for the model based on the data of the other 94 glaciers. Thus, the site-independent GBR is independent from any data of the glacier where the model is tested (see Fig. 2).

The second GBR model is season-independent and aims at "extrapolating" the B_w data in time only (filling data gaps over glaciers with discontinuous records of B_w). For the season-independent GBR, we built a model for each year (when the B_w data is available) of each glacier. In this case, each year of each glacier is used, in turn, as an independent test for the GBR, which is trained using the data of the other years (of the tested glacier) and of the other 94 glaciers.

Moreover, the motivation of including also the data of the other 94 glaciers in the training of the season-independent GBR has been described at lines 217-221:

Thus, unlike the site-independent GBR, the season-independent GBR is informed with data of the glacier where the model is tested, excluding only data of the year when the model is tested. For the season-independent GBR, data of the other 94 glaciers is still used in the training because many glaciers only have a small number of years with available B_w data. Thus, in case of limited B_w data, this may help the GBR to learn from the data of the other 94 glaciers.

- L206-210: this paragraph is very unclear. It's also not clear what the non-GBR specialist can learn from table 1? Either discuss to explain the value of this information or delete.

We added an explanatory sentence at lines 199:204, where the hyperparameters are introduced:

Different hyperparameters characterize a GBR. In this study, we applied a grid search to optimize the number of estimators (number of additive trees, i.e. number of iterations), the maximum depth that each tree can reach, the minimum number of samples required to be at a leaf node of a tree, and the learning rate. The higher the number of estimators or the maximum depth is, the more complex and less generalized the GBR model is. In contrast, the larger the minimum number of samples is, the less complex and more generalized the GBR model is. A smaller learning rate makes the GBR more robust to the specific characteristics of each individual tree, thus allowing a better generalization. However, the smaller the learning rate is, the more subsequent trees (iterations) are generally required to reach the minimum of the cost function. A 10-fold cross-validation was applied with different combinations of hyperparameters. The hyperparameters that were able to minimize the mean squared error of the validation data were chosen. Finally, the GBR model with the chosen hyperparameters was tested on independent data.

Similar information is now reported in the caption of the table. Furthermore, we related the contents of the table with its description at lines 212:226, and we report the reader to the related discussion:

The average optimal hyperparameters for all the studied glaciers are reported in Tab. 1. The resulting site-independent model is more generalized (smaller number of estimators than the season-independent GBR and higher minimum number of samples per leaf), while the season-independent model is more detailed and can be split into individual sub-models adapted to a small number of samples. The different architecture between the site-independent and the season-independent GBRs is discussed in Sec. 5.2.1.

- Section 4.1: intuitively, I would put this section later in the paper. But I leave this open.

Thanks for the suggestion that we considered. We decided to move this section (new Sec. 4.2) after the main results of the study (new Sec. 4.1).

- L250: I don't think that the calendar year should be part of the predictor pool. If a constant line has predictive power, it's because the training data have trends that are not in the reanalysis data, and I think it is highly problematic to rely on such information when trying to extrapolate in space and time. Happy to be convinced otherwise though.

By following also the suggestion of Referee #2, we further discussed the implication of the use of the year in the set of predictors (lines 425-435). We also discussed its limitations for a potential use of the GBRs for extrapolation in the future, which indeed would be problematic (lines 461-466).

- L253-256: this is very unclear, I'm sorry but I don't understand what this means.

We reformulated the sentence in the middle of this paragraph, which we agree was confusing (lines 352:357):

However, skipping reanalysis variables from the set of predictors leads to higher errors for some individual glaciers, especially in the representation of the temporal variability of the B_w data ~~snow~~ accumulation. In fact, excluding the reanalysis variables, the year is the only predictor able to vehiculate the climatological information, in other words, the year is the only predictor that could be used to predict a different adjustment factor depending on the accumulation season (all the other predictors are constant in time).

- L278-281: Isn't this information already on the figure and does it need repeating here?

In the previous round of review, we were asked to add information on the mean absolute errors (which is not on the figure). Despite we agree that there is some redundant information here, we also think that the reader can benefit from the quantitative information provided here, as quantitative values are rarely reported in the text of the manuscript.

- Fig. 5 is very difficult to read.

We agree that a lot of information is provided in the figure. However, it allows having an overall impression of the adjustment factors for the four regions, with the possibility to focus on specific glaciers or to detect outlier glaciers in the regions.

- Fig. 6 illustrates well what is confusing me: why do the correction factors have trends? I think that Fig. 6 would also be a good opportunity to show actual data instead of correction factors, which is a very abstract notion for glaciologists...

Thank you for this comment that we considered. However, we think that there are several reasons to prefer showing biases instead (factors between the winter mass balance and the models' estimates):

- *The goal of the paper is to reduce the biases between the reanalysis and the winter mass balance data.*
- *Showing biases with factors allows comparing the results between different regions receiving different amounts of precipitation. Thus, we can understand in which regions the reanalyses are more biased.*
- *For a similar reason, the figure looks much cleaner by showing factors instead of the winter mass balance (or precipitation) data. In fact, there is a large variability between the winter mass balance data over the years (also because different glaciers are considered over the years depending on the data availability). In contrast, the factors between the winter mass balance data and the reanalysis precipitation (or GBR models) remains much more constant over the years.*

Here, we report the two figures that confirm what we stated above:

Fig. 6 (new Fig. 5) in the manuscript (factors):

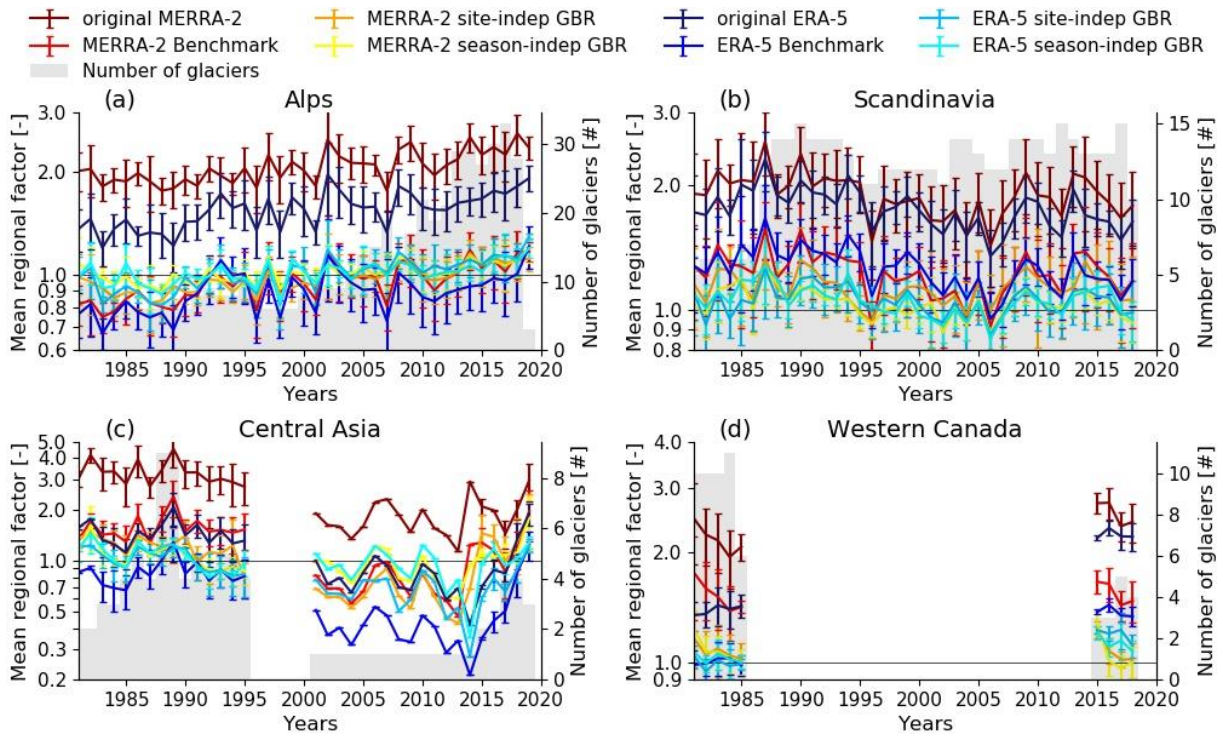
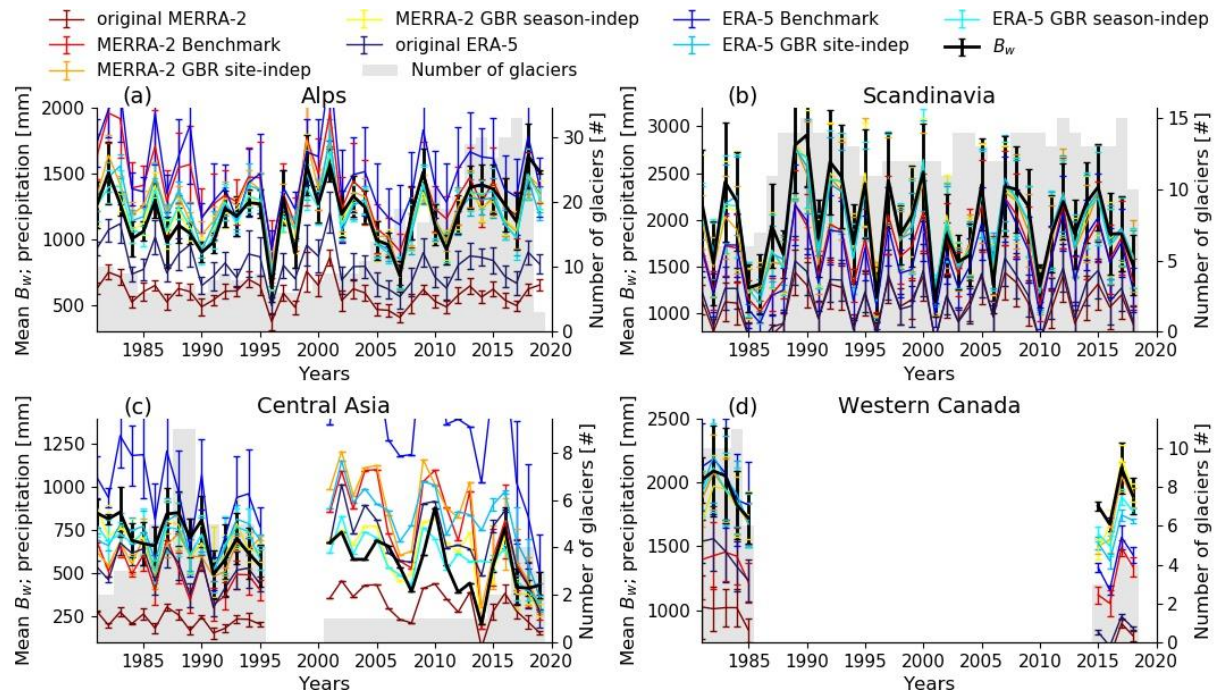


Figure with winter mass balance (or precipitation) data (in the figure above, we basically report the factor between the black line and the other lines (i.e. we report the biases)):



Regarding the first part of your comment, the biases (factors in this case) can have trends. This has been discussed in Sec. 5.2.1 (lines 425:428):

In general, the year was included in the set of predictors for the GBRs because it may allow learning the climatological information and potential trends in terms of reanalysis biases against the B_w data. This can be relevant in case reanalysis variables included in the predictors are not able to represent

the hypothetical trends of these biases, which can exist due to the increasing availability of observational data that could be assimilated by reanalysis models over the years.

- Fig. 6: when averaging factors, you should also plot the range (std dev) to show the robustness of the differences

We agree, the standard deviation has been added to Fig. 6 (new Fig. 5).

- L307: "confirming the importance of a specific optimization scheme depending on the goal of the model." -> I have to reiterate: what is the goal of the model?

This has been made clear by following your previous suggestions.

- L320: genuine question: is there any skill in ingesting data from very far away glaciers to interpolate in time?

Thanks for your comment. We think that the answer is provided by Fig. 6 (b, d, f and h) in the new manuscript, and is discussed in Sec. 5.2.2.

- L321: "In conclusion, filling data gaps is much simpler than estimating SWE on glaciers with no observations." yes, and this raises the question whether GBR is really needed for that or not (rhetorical question, requiring no change to the manuscript)

We agree and we understand your comment. However, we think that it is fair to use the same kind of model (i.e. a GBR) for the interpretation of the differences between site- and season-independent models.

- L325: see comment above: it is really unclear to me how the profiles are predicted...

As for the comment above, we think that after the changes in the manuscript related to your comment above, this is now clear to the reader.

- L363: "This suggests that complex models such as our GBRs are needed to adjust reanalysis to different glacier sites" -> this statement is made based on the benchmark model, which is not data informed. The paper does not say which model complexity is needed to achieve WMB reconstruction.

Thank you. We modified the sentence by specifying what you correctly commented (line 378-380).

- L380: "A disadvantage of tree-based algorithms, however, could be that this approach does not predict continuous values." -> I feel that this information should be shared much much earlier in the paper.

Thank you. This information is now provided in the description of the model in the Methods section (3.2.2, lines 186:187).

- Section 5.1.2 is highly speculative, short and not convincing.

We agree. We added another reason that, intuitively, made us preferring a tree-based algorithm to other machine-learning methods, and that we think consolidates our choice. Here, we report the new section (lines 394:405):

Intuitively, we preferred a tree-based algorithm given the high inhomogeneity in terms of spatial distribution of the considered glaciers. As further discussed in Sec. 5.2.1, a tree-based algorithm can exploit the coordinates (if provided as predictors) to easily split into individual sub-models adapted to different regions of the world (from the continental scale to the glacier-specific scale). Such operations would not be possible by using a simpler model such as a multiple linear regression. Also, it is less clear to us how artificial neural networks would behave given the considerable inhomogeneity of the spatial distribution.

[In fact](#), we have chosen a tree-based algorithm because of its higher readability in terms of the predictors' usage compared to other machine-learning methods (e.g. Huysmans et al., 2011; Freitas, 2014). A disadvantage of tree-based algorithms, however, could be that this approach does not predict continuous values. Yet here, we aim at predicting an adjustment factor depending on a classification based on the used predictors, which is exactly the purpose of a tree-based algorithm. The choice of a gradient boosting instead of other tree-based algorithms (e.g. random forest (Breiman, 2001)) is motivated by the fact that gradient boosting is a gradient descent algorithm, where each additional tree tries to reduce the bias (which is the main goal of our study) rather than the variance of the predictions.

- In general, the discussion is by far the most interesting part of the paper. Many points related to how the method works or doesn't work are described here, and this is what makes the paper interesting.

Thank you.

- Section 5.2.4: I might be wrong, but I think that this is the only time the difference between the two reanalysis datasets is discussed? Does this justify the additional complexity of many of the figures? (I'm not suggesting changing the study design at this stage, but it is still a valid question).

It is true that the differences are not discussed in depth and we put more emphasis on the methodology, but we think that the provided information can be useful for other scientists, e.g. for the calibration of models that need an adjustment of reanalysis precipitation. In fact, from the paper the reader can also have an idea on precipitation biases (of reanalysis datasets among the most used worldwide) for glaciers in different regions of the world. Also, we believe that showing that the model provides similar results with two different reanalyses, makes the study more robust.