Response letter

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

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

Dear Editor,

Many thanks for your valuable work for the sake of science. We are also grateful to the reviewers' time and efforts to help improving our study.

In response to the two reviews we have thoroughly and substantially revised our manuscript. Below, please find a brief summary of the most important changes of the study (incl. substantial re-calculations, adjustments of the analyses schemes and a new sensitivity study) and related changes in the manuscript:

- The manuscript has been shortened by removing the SWE-trends analysis. Thus, the general goal, the abstract, the introduction and the conclusion have been rewritten by focusing more on the gradient boosting regressor (GBR) models and their ability of adjusting the reanalysis.
- The cross-validation and test schemes have been modified. Each glacier, in turn, is used to test the model trained and validated with the other glaciers. All the results concerning the GBR models have thus slightly changed. However, none of the changes affected our discussion and the conclusion.
- The analysis of the predictors' importance has been performed differently. The overall GBR models' performance in terms of root-mean-squared-error (against the snow accumulation data) is evaluated depending on different groups of predictors.
- The spatial generalization capability of the models is now discussed in a more critical way.
- The ability of the models to represent the temporal variability of the snow accumulation has also been more critically discussed and a new sensitivity study is reported in the supplement.

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 the Editor

Legend: Editor's comments; *authors replies*

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

Some minor comments after my overview read are:

(a) when you review the state of the art in the introduction, I think it would be good to at least mention dynamical downscaling, which has been employed more and more over glacier regions in recent years. A relevant paper could be Mölg & Kaser (2011, JGR Atmospheres, 116: D16101). I want to emphasize that I am not suggesting this source because I am an author on it, but because if it was one of the first (or the first) study to evaluate high-res dynamical downscaling against measurements taken on a glacier (including precipitation). Hence the study could be a relevant addition to the refs:

Dynamical downscaling and Mölg & Kaser (2011, JGR Atmospheres, 116: D16101) are now mentioned in the introduction at lines 58-60:

~~In fact, the~~ As a result, downscaling of precipitation estimates of reanalyses is necessary to represent the local conditions in high-mountain regions. Different statistical and dynamical downscaling methods exist (cf. Maraun et al., 2010), which has also been employed and evaluated over glacierized regions (e.g. Mölg and Kaser, 2011).

(b) The very large tables 2 and 3 could be moved to the supplement:

Tab. 2 was moved to the supplement, while tab. 3 was shortened by including only glaciers with a record of at least 15 season of snow accumulation data. This table was kept in the manuscript because it is needed to support the analysis of the models' performance to represent the temporal variability of the snow accumulation.

(c) I am also saying this because I would avoid having a supplement for only one figure:

The supplement is now composed by three figures and a table.

Comments by Reviewer 1

Legend: Reviewer's comments; *authors replies*

Manuscript excerpts with <u>added text</u> and ~~deleted text~~

We would like to acknowledge the reviewer for this thorough and critical review that has helped us to sharpen the focus of our study.

The problems begin with the title, which overstates its importance. Only a tiny fraction, in fewer than half of the continents, of the world's glaciers are examined. The manuscript has too many figures and tables. The manuscript is supposed to be within 12 journal pages for TCD. The tables and figures alone, most of which occupy a full page, would take up this much space. The figures are bloated. For example, there is no need to illustrate "Tree 1" nor "Tree N", both of which are identical in Figure 3. The PCA section (4.1) doesn't tell the reader much more than the fact that elevation is the most important downscaling predictor:

Title:
We agree that the term "world's glaciers" can be misleading. In response to this comment, we changed the title to: "Snow accumulation over glaciers in the Alps, Scandinavia, Central Asia and Western Canada (1981-2020) inferred from climate reanalysis and machine learning"

Number of figures and tables:
We agree that some simplification was beneficial to the paper and we accordingly performed major changes including a reduction of the number of Figures / Tables:

- *The manuscript was shortened by removing the SWE-trends analysis (old Sec. 3.4, 4.3 and 5.2; Tab. 4; Fig. 11 and 12 were removed). Thus, the general goal, the introduction and the conclusion were modified by focusing more on the GBR models and their ability of adjusting the reanalysis.*
- *Tab. 2 was moved to the supplement.*
- *Tab. 3 (new Tab. 2) was simplified by including only glaciers with a record of at least 15 season of snow accumulation data. This table was kept in the manuscript because it is needed to support the analysis of the models' performance to represent the temporal variability of the snow accumulation.*
- *Fig. 2 was moved to the Supplement.*
- *Fig. 3 was simplified and replaced by a smaller figure without the illustration of the "Trees" (new Fig. 2), which only described the cross-validation and test schemes.*
- *We also agree that Sec. 4.1 needed to be rewritten in order to better quantify the added value of each group of predictors on the model's performance. Thus, Fig. 4 was removed and Fig. 5 was moved to the supplement as it shows that other predictors than elevation are important to explain different biases between reanalysis' precipitation and snow accumulation on glaciers. In the revised version of the paper, we showed the changes in terms of overall model performance when suppressing groups of predictors (new Fig. 3).*

The results of the new Fig. 3 are described at the beginning of Sec. 4.1:

<u>In order to understand the importance of the predictors used by the GBR models (i.e. those not related to the elevation of the glaciers and their elevation difference with the reanalysis' grid), we evaluated the changes in terms of overall GBR model performance when suppressing groups of predictors. For both ERA-5 and MERRA-2 site-independent GBR models, the smallest RMSE results when using all predictors (Fig. 3a and b). The RMSE particularly increases when suppressing the MERRA-2 single level and pressure levels variables from the predictors. In turn, for both ERA-5 and MERRA-2 season-independent GBR models, the smallest RMSE results when suppressing the single level and pressure</u>

levels variables from the predictors (Fig. 3c and d). The RMSE increases most when suppressing the year, the topographical parameters and the glaciers coordinates simultaneously as predictors.

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 snow accumulation. In fact, excluding the reanalysis variables, 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). Therefore, and to allow a fairer comparison between site-independent and season-independent GBRs, in all our following analyses we always include all predictors.

The leave one out validation is problematic as there is no independent validation dataset used, meaning that biases in precipitation are unlikely to be identified:

Many thanks for this thought. However, we do not fully agree with this statement. For the "site-independent GBR", the model was always validated on a glacier that was independent from the model's training. Thus, as stated in the manuscript, the leave-one glacier-out cross-validation allowed evaluating the generalization of the machine-learning models for glaciers located in the same regions of the training data. Fig. 7 (old Fig. 9) shows a more robust validation, where the performance of the machine-learning models is also evaluated for completely independent regions (removing neighboring glaciers from the training data). Biases of reanalysis's precipitation against snow accumulation data (based on ground measurements and extrapolation techniques (see Sec. 2.2)) on the glaciers of the study are therefore identified (see Fig. 4 and 5 (old Fig. 6 and 7)).

Despite the glaciers used for validation being independent from the GBR model's training, it is true that they had an influence on the choice of the optimal hyperparameters of the GBR model, i.e.: the GBR model was optimized to perform well on the validation data. However, each single glacier (1 out of 95 glaciers) used for the validation had a very limited weight on the overall performance (mean squared error) and on the choice of the GBR's hyperparameters.

In order to make the proposed methodology even more robust, we also defined the hyperparameters independently from the test sites, i.e.: in turn, each glacier was used to test the GBR model trained and validated (10-fold cross-validation for the selection of the hyperparameters) with the other glaciers. As a consequence of the new training, all the results regarding the GBR models have slightly changed. However, this did not affect our discussion nor the conclusion of the study.

The new cross-validation and test scheme is illustrated in the new Fig. 2. and is described at lines 191-210:

Different hyperparameters characterize a GBR. In this study, we applied a grid search to optimize the number of estimators (number of additive trees), 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 ~~maximum number of predictors that are randomly selected at each split for each tree (the predictor reducing the error the most is used to split the node)~~ learning rate. 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. ~~The optimal values are reported in Table 1.~~ Finally, the GBR model with the chosen hyperparameters was tested on independent data.

The validation ~~data was~~ 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 ~~schemes (see Fig. 2). The first one~~ and test schemes. The first GBR model is site-independent and aims at "extrapolating" the Bw data in time and space (over glaciers with no Bw data). ~~We thus applied a leave-one-glacier-out~~ Thus, groups of data (folds) in the 10-fold cross-validation~~. The second one~~ contain data of different glaciers and the site-independent GBR model with the chosen hyperparameters was tested on an independent glacier. This process was repeated for each glacier, which was used to test the GBR-model defined with the data of the other glaciers (see Fig. 2). The second GBR model is season independent and aims at "extrapolating" the Bw data in time only (filling data gaps over glaciers with discontinuous records of

Bw). For these cases, ~~we applied a leave-one-season-out~~ groups of data in the 10-fold cross-validation ~~. Thus, the~~ contain data of different years but different groups can contain data of different years of the same glacier. Finally, the season-independent GBR model with the chosen hyperparameters was tested for an independent year of a given glacier. This process was repeated for each year and each glacier.

The average optimal hyperparameters for all the studied glaciers are reported in Tab. 1. The resulting site-independent model is more generalized (since no information regarding the glacier where the model is validated and tested was provided), while the season-independent model is more detailed and perf~~orms better over glaciers with~~ can split into individual sub-models adapted to a small number of samples (since it can exploit the Bw data ~~but worse over glaciers with no Bw data.~~ of the tested glacier).

ERA-5 and MERRA-2 reanalyses are used without any mention of their potential large biases in the mountains. For example, Liu and Margulis (2019) report that MERRA-2 underestimates snowfall (which is based on the "PRECTTOLAND" variable used here) by 54% in High Mountain Asia.:

We are fully aware of the limitations of Reanalyses (because of missing and/or highly inaccurate in-situ observations) in high mountain region and specifically precipitation. In fact, our whole study is in principle motivated by this major challenge of improving the quantification of high altitude (solid) precipitation and SWE. In the manuscript, reanalysis biases in high-mountain regions were thus mentioned including references in the introduction. However, we agree that the biases observed in previous studies have not been described and quantified abundantly enough. In the revised paper we better included them in the introduction thus enhancing the comprehensiveness of the manuscript. We also added respective reference in the revised manuscript (lines 53-57).

However, the performance of reanalysis results can vary greatly depending on the region and the elevation range of interest (Sun et al., 2018). Large biases in reanalysis precipitation are particularly observed in high-mountain regions (e.g. Liu and Margulis, 2019; Zandler et al., 2019) . The scarcity of observations ~~that can be assimilated~~ available for assimilation and the coarse resolution of such models limit their accuracy in areas of complex topography and their suitability for studies at a local scale (e.g. Salzmann and Mearns, 2012, (for snow)).

New cited references:

- *Zandler, H., Haag, I., and Samimi, C.: Evaluation needs and temporal performance differences of gridded precipitation products in peripheral mountain regions, Scientific Reports, 9, 15 118, https://doi.org/10.1038/s41598-019-51666-z, 2019.*
- *Liu, Y. and Margulis, S. A.: Deriving Bias and Uncertainty in MERRA-2 Snowfall Precipitation Over High Mountain Asia, Frontiers in Earth Science, 7, 280, https://doi.org/10.3389/feart.2019.00280, 2019.*

It's not clear to me that the downscaling techniques presented here will correct that bias, as no independent evaluation of precipitation is presented:

Reanalysis's precipitation is compared against snow accumulation data on glaciers. This data clearly is independent, and it is to our knowledge the only and thus best possible source of (cumulative) precipitation at very high elevation. The machine-learning model is trained, validated and tested against these snow accumulation data on glaciers. In general, from the results presented in the revised manuscript (e.g. Figs. 4 and 5) it is clear that, on average, the machine-learning models can adjust the reanalysis' bias against snow accumulation on glaciers, which is among the main purposes of the study. We hope that this response answers the reviewer's comment. Otherwise, we would be happy to obtain additional explanations.

Melt and sublimation are ignored in the "winter mass balance," which is then the wrong term:

We do not fully agree with the reviewer here. The term "winter mass balance" refers to the snow water equivalent found on the glacier close to the maximum of snow depth, or the end of winter. Therefore, the winter mass balance – per definition – includes loss terms such as melt and sublimation, although they are not individually quantified. Furthermore, our periods of analysis are adjusted to optimally match the period where the components of melt and sublimation are small in comparison to accumulation by solid precipitation.

After carefully searching through the text, I still cannot understand how precipitation phase was treated. It seems to have been ignored as SWE is used interchangeably with the downscaled precipitation on glaciers. But then, in Table B1 and B2 ERA-5/MERRA-2 snowfall variables are listed as predictors?:

Indeed, the precipitation phase was ignored. In the revised paper (lines 144-152) we more clearly described this choice and the reason of including the snowfall variable in the predictors:

First, we derived total or average of all variables provided by the reanalyses for the entire accumulation season. Subsequently, a machine learning model to adjust the total precipitation (see Sec. 2.1.1 and 2.1.2) of the reanalyses over glaciers for the accumulation season was developed to derive SWE estimates. We use a GBR (gradient boosting regressor), which makes use of several meteorological variables (original and downscaled) and topographical parameters as input variables (predictors). In principle, a different adjustment factor of precipitation might be needed depending on the precipitation phase. However, as we only adjust the total precipitation occurring during the accumulation season, the adjustment factors used here represent the "average" adjustment factor of all precipitation events. Moreover, the snowfall variable was used as a predictor in order to enable the GBR model to learn that a different "average" adjustment factor must be applied depending on the fraction of snowfall and total precipitation (i.e. depending on the main precipitation phase during the accumulation season).

Comments by Reviewer 2

Legend: Reviewer's comments; *authors replies*

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

Guidicelli et al propose an interesting method to downscale and bias-correct reanalysis precipitation data to the elevation and sites of glaciers in 4 regions of the world. 2 reanalyses are used : ERA5 and MERRA2. The method is based on gradient boosting regressions, a technique from the field of artificial intelligence. The performance of this method is evaluated through cross-validation and discussed in terms of both temporal and spatial extrapolation. Finally, precipitation trends on glaciers are derived for each 4 regions based on the bias corrected and downscaled reanalysis data.

The study tackles the very interesting and yet unsolved issue of high-altitude precipitation amounts, with tools from machine learning. It adds to the existing literature by focusing on glacier winter mass balances, used as a proxy for winter precipitation at high altitudes. In my opinion, this makes the topic of this study very relevant. While the analyses displayed are in general sound, I advise a revision of the paper with respect to concerns regarding the spatial generalization capability of the models and the derivation of trends, see below.

We would like to thank the reviewer for the positive appreciation of our work and the constructive comments that have helped us to improve the paper considerably.

MAIN COMMENTS

1 - Comparison/justification with respect to other AI techniques for bias correction and downscaling in literature : Even though the introduction describes well the existing literature on AI-based downscaling/bias correction methods, the choice of GBR is barely justified with respect to other techniques. I would have expected elements in that direction in the manuscript, especially since a section of the Discussion is entitled : '5.1 Advantages and disadvantages of gradient boosting regressors':

The discussion section was reorganized and a subsection (5.1.2) was dedicated to explain our choice of applying a gradient boosting regressor model (lines 378-385):

5.1.2 Differences with other machine learning algorithms 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.

New cited references:

- *Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.*
- *Freitas, A.: Comprehensible classification models: a position paper, ACM SIGKDD explorations newsletter, 15(1), 1–10, 2014.*
- *Friedman, J. H.: Greedy function approximation: a gradient boosting machine, Annals of statistics, pp. 1189–1232, 2001. (this paper is cited in the introduction (line ?))*
- *Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., and Baesens, B.: An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models, Decision Support Systems, 51(1), 141–154, https://doi.org/10.1016/j.dss.2010.12.003, 2011.*

2 - Limits inherent to the number of available learning data :

Some of the regions of interest, e.g. Canada and Central Asia, have in total less than 20 glaciers used in this study, which is an extremely low percentage of the number of glaciers that they truly host.

This in my opinion strongly impedes the (spatial) generalization capability of the GBR models learned on these data, to the region of interest as a whole. Although this is not what the authors do in the paper, this is what the title suggests while mentioning the world's glaciers. I would strongly recommend to modify this misleading title, as the developed technique is in practice not applied to derive precipitation data over any glacier of the world, but is limited to (i) the regions of interest and (ii) the few glaciers with data in these regions..

On top of the low sampling level for application of machine learning techniques in general, there may be furthermore a strong sampling bias in the glaciers data from WGMS, for instance towards large glaciers in the European Alps, so that the representativity of the glaciers with data w/r to the regions of interest is questionable. It follows that it is hard to know whether models or conclusions inferred solely based on these very few glaciers, are representative of the region as a whole.

I very much would like the authors to comment on this.

"The good performance of the GBRs in terms of bias suggests that they can be used for SWE estimates over glaciers where no ground observations are available (site-independent GBRs)". Despite being better than the benchmark, the performance of site-independant GBR models is limited (Fig 9) and decreases when data of neighbouring glaciers are excluded from the training. Considering that, and the likely sampling biases of WGMS data, I think the authors could revise this sentence:

We agree with the reviewer regarding most aspects mentioned here. In the revised paper we discussed more critically our approaches and also demonstrated the limitations of our approach, for example in the case of a limited number of observations.

Title: We agree that the term "world's glaciers" can be misleading. We changed the title to: "Snow accumulation over glaciers in the Alps, Scandinavia, Central Asia and Western Canada (1981-2020) inferred from climate reanalysis and machine learning"

Regarding the sentence mentioned ("The good performance of the GBRs in terms of bias suggests that they can be used for SWE estimates over glaciers where no ground observations are available (site-independent GBRs)") we fully agree that our statement was too optimistic / too general. This was better specified at lines 317-319:

However, the performance generally decreases when the glacier is not in proximity to the glaciers used to train the GBR models. Furthermore, we assume that the resulting performance strongly depends on the characteristics of the glacier with respect to the glaciers used in the training.

and in the discussion at lines 375-377:

Both, the GBR models and the benchmark do not require direct ground observations to be applied. However, the performance of the GBR models is influenced by the amount of data used to train the models and strongly depends on the characteristics of the glacier with respect to the glaciers used to train the models.

"3 - Trends :
In my opinion the derivation of trends based on the GBR modelled precipitation, should be accompanied with sensitivity tests to ascertain the robustness and uncertainties of this method. Typically, data-withdrawal techniques could be used on the longest time-series to evaluate the robustness/uncertainty of the trends derived when missing data are encountered. The distribution of the data gaps within the time-series (= for instance one missing season every two year, vs 20 years with data and nothing for the following 20 years) may also play a role, and it would be good to have an insight into this and possibly only derive trends for glaciers with a sufficient number data (seasons). The strong limitation of temporal extrapolation for some glaciers is highlighted l 350-l355, hence making a derivation of trends on these glaciers meaningless."

Thanks a lot, this is a very valid comment and a good suggestion.

In the trend analysis, the GBR models were applied over 41 years for all the glaciers of the study. The Bw data was only used to train the GBR models and not to derive the trends. We used the temporal correlation (over years) between the GBR models and the Bw data as an indicator of the trends accuracy. However, as highlighted, the number of glaciers with long records of Bw data is limited and do not allow general conclusions in terms of trends. For this reason, and in order to reduce the length and sharpen the focus of the manuscript, we decided to completely remove the trend analysis from the manuscript.

Nevertheless, we still discuss the possibility and the limits of deriving trends with our GBR models. This was supported by a new sensitivity test as proposed by the reviewer: the temporal correlation of the season-independent GBRs with the Bw data was evaluated depending on the number of years of data of the tested glacier used to train the GBR model (similarly to Fig. 7a, c, e and g (old Fig. 9)). This sensitivity test was performed only for glaciers with more than 30 years of Bw data available and the result is reported in the supplement.

In the first version of the manuscript, the trends were also derived with the site-independent GBRs, which are not affected by the number of years with available Bw data (because no Bw data of the tested glacier is used for training). Tab. 2 and Fig. S3 show that the site-independent GBRs often perform better than the season-independent GBRs in terms of temporal correlation with the Bw data. This indicates that the number of available years with Bw data does not necessarily need to be high in order to accurately represent the temporal variability of the snow accumulation over the years and thus, in order to derive trends.

The new results are described in Sec. 4.2.3 and a new dedicated section is reported in the discussion (5.2.3):

5.2.3 Representation of the temporal variability of the snow accumulation

All GBRs aimed at minimizing the MSE between the predicted and reference logarithmic adjustment factors (Eq. 3). The improvement of the temporal correlation between the original reanalysis and the Bw data is thus a consequence of bias-adjusted estimates over accumulation seasons rather than a primary goal of the GBRs. A sensitivity test reported in the Supplementary material (Fig. S3) suggests that the season-independent GBRs are not very sensitive to the number of years of data of the tested glacier used for training. Their performance is comparable to the site-independent GBRs (Tab. 2). Furthermore, only in a few cases the site-independent GBRs show a performance inferior to the original reanalysis or the benchmark method (e.g. Ts. Tuyuksuyskiy glacier). These promising results suggest that our new estimates could also be used to derive SWE trends with generally higher accuracy than the original reanalyses, thus potentially providing insights on the relation between climate change and both snow accumulation and precipitation at the highest elevations of mountain ranges, where virtually no direct precipitation records are available. Still, the limited number of glaciers with abundant Bw data coverage available over sufficient number of years do not allow us to perform a complete application of this approach.

MINOR COMMENTS

- the GBR consider as predictors both elevation differences between reanalysis pixel and glacier site, and downscaled variables like temperature, whereby the downscaling of temperature itself mostly relies on this altitude difference. Hence there is a high redundancy in the chosen predictors. Did you test suppressing the downscaled predictors ?

Thanks for this interesting comment. The high correlation between predictors is a problem for the interpretability of the predictors' importance. However, this does not importantly affect the performance of the GBR because decision trees are by nature not affected by multi-collinearity. If two predictors are highly correlated, the tree will choose only one of the two predictors when deciding upon a split.

However, we agree that Sec. 4.1 needed to be modified in order to better quantify the added value of each group of predictors on the model's performance. Thus, Fig. 4 was removed and Fig. 5 was moved to the supplement. In the revised version of the paper we showed the changes in terms of overall model performance when suppressing groups of predictors (new Fig. 3). The results of the new Fig. 3 are described at the beginning of Sec. 4.1:

In order to understand the importance of the predictors used by the GBR models (i.e. those not related to the elevation of the glaciers and their elevation difference with the reanalysis' grid), we evaluated the changes in terms of overall GBR model performance when suppressing groups of predictors. For both ERA-5 and MERRA-2 site-independent GBR models, the smallest RMSE results when using all predictors (Fig. 3a and b). The RMSE particularly increases when suppressing the MERRA-2 single level and pressure levels variables from the predictors. In turn, for both ERA-5 and MERRA-2 season-independent GBR models, the smallest RMSE results when suppressing the single level and pressure levels variables from the predictors (Fig. 3c and d). The RMSE increases most when suppressing the year, the topographical parameters and the glaciers coordinates simultaneously as predictors.

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 snow accumulation. In fact, excluding the reanalysis variables, 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). Therefore, and to allow a fairer comparison between site-independent and season-independent GBRs, in all our following analyses we always include all predictors.

- the predictors in the PCA figures (4 and 5) are often barely lisible. Fig 5 could maybe join the supplemental material.

Fig. 5 was moved to the supplement. We also increased the fontsize and avoided the overlapping of predictors' names.

- l 264-274 : could the different magnitude in factors relate to known biases / weaknesses of the reanalyses in representing different types of precipitation events ?

Yes, thanks for this good suggestion. However, so far we were not able to directly relate our findings to statements reported in the literature.

- l 311 : "their performance is worse than the site-independent models". It is not so clear for me why : could you please explain ?

The season-independent GBR model has a higher number of trees and less samples are needed to create a new leaf of the tree (i.e. to predict a different adjustment factor) than the site-independent GBR. Thanks to its higher complexity than the site-independent model, if Bw data of the validated glacier is used to train the season-independent model, this latter can learn the specific characteristics of the tested glacier and perform better than the site-independent model. On the other hand, if no Bw data of the tested glacier is used to train the season-independent GBR, its performance is worse than the site-independent GBR, because it will overfit the training data.

A new dedicated section is reported in the discussion:

5.2.1 Site-independent and season-independent GBRs

The lower generalization of the season-independent GBRs compared with the site-independent GBRs allows the splitting into individual sub-models adapted to a small number of samples (see Tab. 1). This enables to exploit the Bw data of the tested glacier by creating a specific sub-model, but can result in an overfit of the training data. In contrary, the higher generalization of the site-independent GBRs allows learning on overall relationships between the used predictors and the reference adjustment factors (Eq. 2).

The used training data and the selected hyperparameters also have a direct influence on the predictors needed by the GBR models to reduce the cost function (Eq. 3). In fact, the use of reanalysis variables (from single level and pressure levels) as predictors, caused an increase of the overall RMSE of both ERA-5 and MERRA-2 season-independent GBRs against the Bw data of all glaciers of the study (Fig. 3c and d). However, despite the high correlation of the downscaled reanalysis variables (cf. Section 3.1) with the elevation of the glaciers, their inclusion in the set of predictors for the training of the site-independent GBRs reduced the overall RMSE (Fig. 3a and b). This difference can be explained by the combined effect of using data of the tested glacier in the training of the season-independent GBRs, and defining a small minimum number of samples required to create a leaf node of the GBR. In fact, the season-independent GBR can theoretically exploit the coordinates to split into individual sub-models adapted to individual glaciers. Therefore, the season-independent GBRs can learn the adjustment factors observed in the other accumulation seasons of the tested glacier and predict a similar adjustment factor for the tested accumulation season, with no need of learning overall relationships between the reanalysis predictors and the reference adjustment factors (Eq. 2).

- l 448 : why were more topographic predictors used in the ERA-5 GBRs than in the MERRA-2 ones ?

We used all the topographical predictors describing the reanalysis's subgrid complexity of both reanalysis products and ERA-5 is providing more descriptors than MERRA-2. This is now specified at lines 450-453:

The differences between the performance of our GBR models are also caused by the different predictors that have been used. For instance, ~~the ERA-5 GBR models use more~~ we considered all the topographical predictors describing the ~~complexity of the grid cell topography than the~~ reanalysis's subgrid complexity of both reanalysis products and ERA-5 is providing more descriptors than MERRA-2 ~~GBR models (see Fig. 3,~~ (see Tab. B1 and B2).

- Fig 2 could join the Supplemental material

Yes, we agree. Fig.2 was moved to the supplement.

- Fig 6 : could the absolute biases also be mentioned ?

Yes, we also evaluated and reported the mean absolute error in addition to the root mean squared error. However, the figure is already too busy to allow more numbers and we reported the results in the text (lines 273-283):

Figure 4 shows the comparison between all glacier-wide Bw values and the models' estimates. MERRA-2 precipitation underestimates Bw more importantly than ERA-5 precipitation in all regions (Fig. 4a and b), with an overall RMSE of 946 mm (mean absolute error (MAE) of 749 mm) against 793 mm (611 mm) of ERA-5. Excluding the Alps, the correlation between the Bw data and the ERA-5 precipitation is always higher than the correlation with the MERRA-2 precipitation. The adjusted estimates obtained with the site-independent and the season-independent GBRs allowed us to consistently reduce (increase) the bias (correlation (r)) between the precipitation of the original reanalyses and Bw (~~from an overall RMSE (CORR) of 946~~ overall RMSE of 433 mm~~(0.74) and 793~~, MAE of 326 mm~~(0.81) of~~ , r of 0.86 for the:MERRA-2 ~~and ERA-5, to 443~~ site-independent GBR; RMSE of 410 mm~~(0.85) and 422~~, MAE of 307 mm~~(0.86) of the~~ , r of 0.87 for the ERA-5 site-independent ~~GBRs, and 287~~ GBR RMSE of 293 mm MAE of 211 mm~~(~~ , r of 0.94 ~~) and 272~~ for the MERRA-2 season-independent GBR RMSE of 275 mm~~(0.95) of the~~ MAE of 200 mm, r of 0.94 for the ERA-5 season-independent ~~GBRs~~ GBR). These results demonstrate the need of an adjustment of reanalyses data to reproduce snow accumulation on glaciers, which are, otherwise, largely underestimated in all four regions involved in this study.

- Fig 7: a ranking of the glaciers with respect to altitude, or to the number of seasons with Bw_data, would enable to more efficiently support the analysis related to this figure, please consider this. The same applies to Fig 11.

Thanks for the suggestion. We modified Fig. 7 (new Fig. 5) by ranking the glaciers with respect to the number of seasons with Bw data . Fig. 11 was removed.

- Tables 1 and 2 could join the supplemental material

Tab.2 was moved to the supplement. In addition, the number of glaciers reported in Tab. 3 (new Tab. 2) was reduced. Tab.1 was kept in order to show the differences in terms of hyperparameters (and generalization) between the site-independent and season-independent GBRs.

- Section 5.2 : this recent literature could also be of interest : https://doi.org/10.5194/hess-24-5355-2020; https://doi.org/10.5194/essd-14-1707-2022 (update of Durand et al., 2009).

Thanks. However, Sec. 5.2 was removed in the revised manuscript.