

The Cryosphere Discuss., author comment AC2
<https://doi.org/10.5194/tc-2022-69-AC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

Matteo Guidicelli et al.

Author comment on "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" by Matteo Guidicelli et al., The Cryosphere Discuss., <https://doi.org/10.5194/tc-2022-69-AC2>, 2022

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

In the following, we report our responses (bold) to the reviewer's concerns (within quotation marks).

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'."

We decided to use a tree-based algorithm because of its higher readability in terms of the predictors' importance compared to other methods (e.g. neural networks). Furthermore, gradient boosting is a gradient descent algorithm, where each additional tree tries to get the model closer to the target and reduce the bias rather than the variance (which is what a random forest algorithm does). We agree, however, that more background on our choice is needed. A dedicated discussion will be added to Sec. 5.1.

"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 will more critically discuss our approaches and also demonstrate 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 will change 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 and we will better specify that the model can be applied on other glaciers, only if the glacier is in proximity to the glaciers used in the training. Moreover, we will specify that the resulting performance strongly depends on the characteristics of the glaciers with respect to the glaciers used in the training.

"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 I 350-I355, 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 are 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. Thus, we propose the following sensitivity test to be included in the revised manuscript: similarly to Fig. 9a, c, e and g and only for glaciers with long timeseries of Bw data, we will show the trends depending on (a) the used number of training seasons for the validated glacier and (b) the

distribution of the available Bw data. The sensitivity test would thus allow us to further evaluate the general expected robustness and uncertainties of the trends depending on the number of years with available Bw data used for training. However, this is only feasible for the season-independent GBR.

In fact, trends are 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 validated glacier is used for training). The fact that the site-independent GBRs often perform better than the season-independent GBRs in terms of temporal correlation with the Bw data, is an indicator 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. In the revised manuscript, we will determine the trends only for glaciers with long time-series of Bw data, i.e.: only for glaciers where the temporal correlation between the GBR models and the Bw data can be evaluated. For these glaciers, the comparison between the trends obtained from the season-independent and site-independent GBRs will allow a better discussion of the potential use of the site-independent GBRs for the derivation of trends on glaciers with no Bw data.

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 only a problem for the interpretability of the predictors' importance. However, this does not 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.

As suggested, in the revised paper, we will show the changes in terms of overall model's performance when suppressing the downscaled predictors (and/or other variables, e.g. topographical) in Sec. 4.1. This will be helpful to quantify the added value of each group of predictors. Correspondingly, Fig. 4 will be modified. In fact, we are quite confident that this will be a better evaluation of the predictors' importance than only showing the frequency of use of the main predictors (Fig. 4a and b).

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

Fig. 5 will be moved to the Supplementary material. We will also increase the fontsize and avoid the overlapping of predictors' names.

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

Yes, this is a good suggestion and we will invest more time in this, trying to link our results with the literature.

- "I 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 validated glacier and perform better than the site-independent model.

On the other hand, if no Bw data of the validated 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.

This discussion will be added in shorter form to the revised paper.

- "I 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.

- "Fig 2 could join the Supplemental material"

Yes, we agree.

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

Yes, we will also evaluate and report the mean bias error in addition to the root mean squared error. However, the figure is already too busy to allow more numbers and we will discuss the results in the text.

- "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 will modify Fig. 7 and Fig. 11 as proposed.

- "Tables 1 and 2 could join the supplemental material"

Yes, we agree.

- "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. We will include this literature in the discussion.