

## Anonymous reviewer 1

---

### 1 Substantial comments

This paper presents a reconstruction of the surface mass balance (SMB) of all French glaciers for the period 1967-2015 based on a deep learning (DL) approach. It strongly relates to the study on the methodology recently published by the same authors in *The Cryosphere*. The data set is comprehensive, interesting and certainly deserves publication. However, there are presently some weaknesses in the presentation of the findings, as well as in the validation of the approach.

---

We are grateful for the overall positive comments by the reviewer. We believe the comments highlighted some aspects that ought to be clearer, and served to further develop some analysis to increase the quality of the paper. All comments have been answered, including the changes made in the manuscript, presented in bold to distinguish them from the unchanged sentences in the updated sections.

Some relevant improvements have been done to the methods during the review process. We have trained a new cross-validation ensemble of 60 members and updated the dataset results. This new ensemble is based on weighted bagging (Hastie et al., 2009) of Leave-Some-Years-and-Glaciers-Out cross-validation (Bolibar et al., 2020), which balances the training data in the model in order to better take into account the lack of data between 1967-1983. The main results and conclusions have not changed, only leading to a slightly less negative average mass balance (from  $-0.72$  to  $-0.71$  m w.e.  $a^{-1}$ ), and slightly higher uncertainties due to the increased presence of underrepresented values of the 1967-1983 period (RMSE: from 0.49 to 0.55 m w.e.  $a^{-1}$  and  $r^2$ : from 0.79 to 0.75). We believe this even more rigorous cross-validation leads to more accurate results and uncertainty estimations.

---

#### 1.1 Structure

The paper needs to be restructured. A Data section is needed – this is presently merged into the Methods. I would also suggest that the study site characteristics are presented in a separate section. At present this is contained in the Results section. The Methods section needs a clearer structure, separating between the actual approach and the uncertainty assessment.

---

The methods section has been restructured following the reviewer’s suggestions. Sect. 2 is now called “Data and methods”, and it includes three sections: 2.1 Data: with a brief introduction of the French Alps and the datasets used and their coverage; 2.2 Methods: with an explanation of the methods used to reconstruct the annual glacier-wide SMB values; and 2.3 Uncertainty assessment: with the analysis of the method’s uncertainties and bias.

---

## 1.2 SMB validation

A major problem of the paper as it stands now is, in my opinion, the lack of validation with independent measurements: The training data set for 32 glaciers is based on “remote-sensing” (Rabatel et al., 2016). As this lies the foundation to the entire study, more effort should be invested to describe this data set (methods, uncertainties). The training data set also contains models and assumptions – the annual SMB of these glaciers has not been directly measured. This needs to be emphasized. Measured (!) information on SMB is available from two sources: (1) the direct glaciological surveys, (2) geodetic surveys.

Although (1) is probably included in the training data set, it should explicitly be shown (e.g. figure with cumulative SMBs) how well the DL approach reproduces the observed SMB series. This would give direct evidence on the performance of the approach, independent from the training data set by Rabatel et al. (2016) that also includes model assumptions.

I was surprised to see that the study does not show any direct validation with data on geodetic mass balance. With the annual resolution of the data set presented here, this would be straight-forward to achieve. For many glaciers in the French Alps, geodetic mass balances over varying time periods are available (see e.g. data base of the WGMS). I assume that some of them have already been used in the set up of the training data set, but probably not all of them. Geodetic mass balances would yield a fantastic way to validate the DL-based estimates of regional variability in mass balances. New remote sensing data sets (e.g. TanDEM-X in combination with SRTM, or ASTER) would also allow computing geodetic mass balances for each individual glacier. I see that deriving such a new geodetic data set is beyond the scope of this study, but it is indispensable to compile the available geodetic survey for French glaciers in this regional assessment in order to gain confidence in the results.

---

There are two main ways to validate the results of a model: comparing them to another independent dataset (e.g., the geodetic MB dataset that the reviewer refers to), and applying cross-validation. The use of cross-validations ensures a true out-of-sample validation, allowing the validation of the full dataset. This presents a substantial advantage when few data are available, as all data can be used both for training/calibration and validation. For the case of spatio-temporal data, this needs to be carefully done, as it was discussed in detail in Bolibar et al. (2020). The reconstructed annual glacier-wide SMB series were not validated against other datasets than the ones mentioned in the paper since the vast majority of available data in the region has already been used for training. The dataset from Rabatel et al.

(2016) is extremely useful in (1) the fact that its uncertainties are very close from uncertainties from the glaciological method ( $0.35 \pm 0.06$  m w.e.  $a^{-1}$ ), and (2) the fact that it is calibrated from geodetic mass balances, meaning that the geodetic data explicitly serve to calibrate and validate the bias.

Regarding the validation against glaciological observations, this has been done as part of the cross-validation in the uncertainty assessment. As we have mentioned in different instances of the “Data and methods” section, all details regarding the methods can be found in Bolibar et al. (2020), which is a purely methodological paper. Since the methods in Bolibar et al. (2020) were based on a case study using the very same 32 glaciers used in this study, it means that the methods and cross-validation results are exactly the same as the ones presented in detail in that paper. It also includes a lot of information regarding the dataset of these 32 glaciers, the performance for each glacier and year, as well as detailed plots with the comparison of simulations and observations (Fig. 6 to 10) . Therefore, our intention with the present paper submitted to ESSD is to apply the methods from Bolibar et al. (2020) in order to generate a regional dataset. Since all the details regarding the methods can be found in a separate paper, here we prefer to focus on the results and the conclusions rather than repeating what has already been presented in detail elsewhere.

On the other hand, we agree that there are a few, independant geodetic mass balances for shorter periods available, which can be used to validate the bias for a sub-period of the reconstructions, but its added value is lower than that of a cross-validation over the entire reconstruction period. However, following the reviewer’s suggestions, we have compared the Pléiades geodetic mass balance data from Berthier et al. (2014) and the ASTER geodetic mass balance data from the newly published Davaze et al. (2020) to our reconstructions. Since these two studies cover different sub-periods, the comparisons have been done separately. Both studies cover only the beginning of the 21<sup>st</sup> century, so the relevance of these bias validation is moderate, as our model has been calibrated to reconstruct SMB for over 50 years, with different climate conditions, especially before and after the summer heatwave of the year 2003.

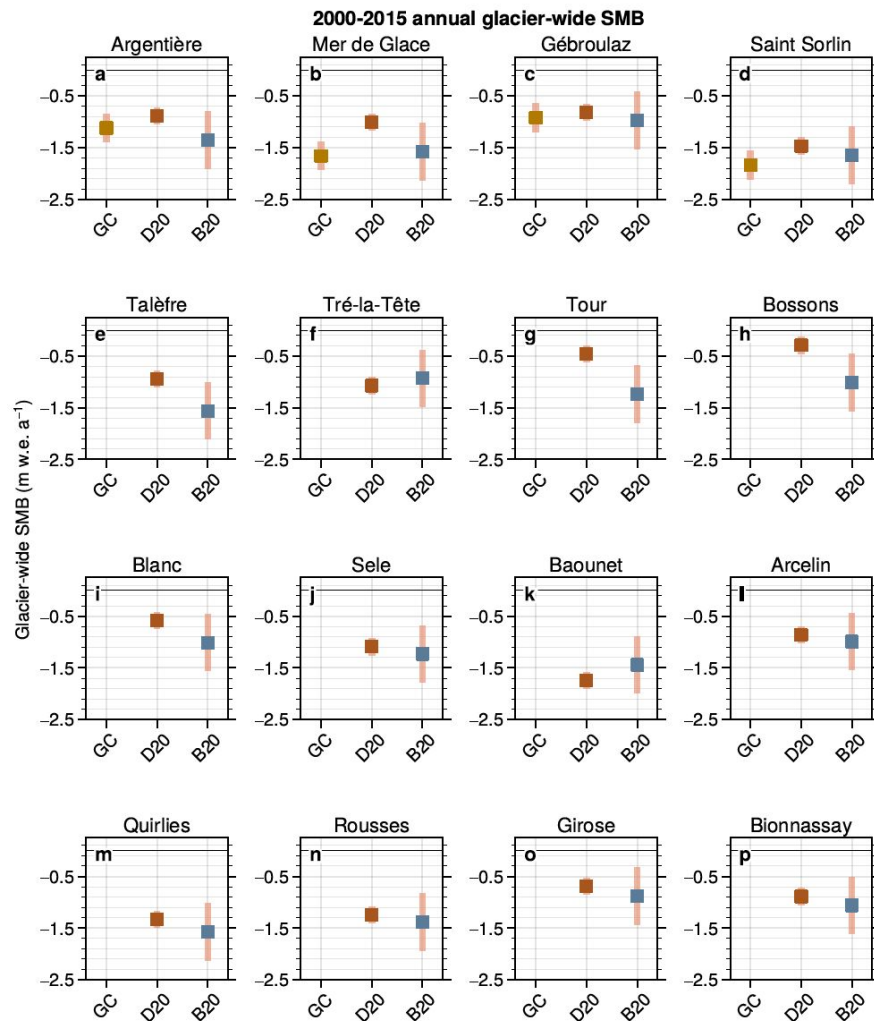
A new section analyzing this (“1 Comparison with independent geodetic mass balance data”), including the two following figures have been added to the supplementary material, in order to illustrate this.

#### “1 Comparison with independent geodetic mass balance data

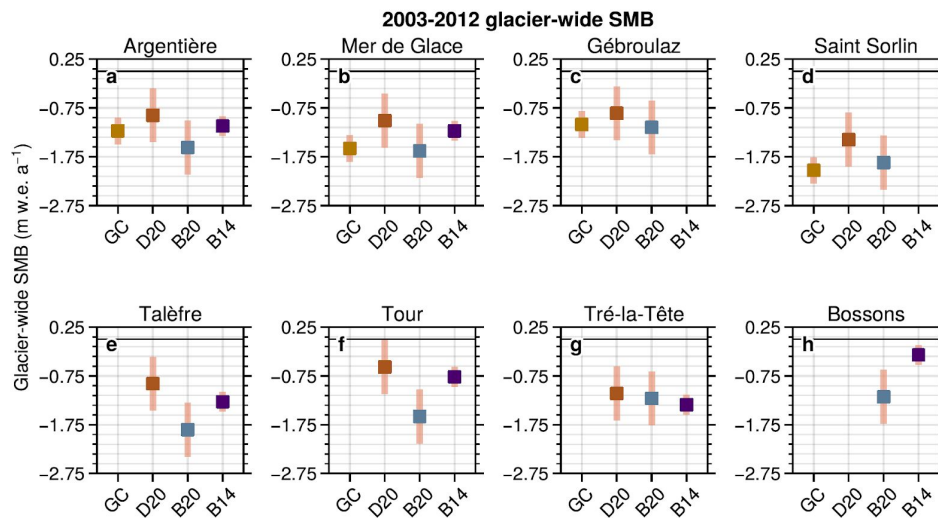
All available annual glacier-wide SMB data in the French Alps have been used to train the SMB ANN of the present study. However, some multi-annual geodetic mass balance (MB) datasets exist that can provide a means to validate the reconstruction’s bias for specific glaciers during multi-annual time intervals. This type of analysis is more limited than the cross-validation done to annual glacier-wide SMB values in Bolibar et al. (2020), as it only gives information about the bias of a sub-period of the reconstructions instead of the accuracy found via cross-validation. Our SMB reconstructions are compared against ASTER

geodetic MB from Davaze et al. (2020) for the 2000-2015 and 2003-2012 periods (Fig. S1 and S2) and against Pléiades geodetic MB from Berthier et al. (2014) for the 2003-2012 period (Fig. S2).

For certain glaciers, the ASTER and Pléiades geodetic MB give slightly less negative MB than the glaciological SMB used to train the deep learning SMB model. This fact might explain the slightly more negative trend of our reconstructions seen for the 2000-2015 and 2003-2012 periods, which experienced very negative SMB after the well known summer 2003 heatwave. This is quite surprising, since both the GLACIOCLIM glaciological SMB measurements and the annual glacier-wide SMB data from Rabatel et al. (2016) have been calibrated with geodetic MB from optically-derived DEMs, which have a very high spatial resolution. Overall, the independent geodetic MB are well within the uncertainty range of our model. There are some exceptions for specific glaciers in the Mont-Blanc massif, such as Bossons, Talèfre and Tour. These glaciers have very large and high altitude accumulation areas, not seen in almost any glacier in our training dataset. On the other hand, for most of the mid-sized glaciers the reconstructions show a good agreement.”



Authors reply to Anonymous reviewer 1 on “A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015”



---

### 1.3 Language

Although the paper is well-written in general, there are several instances where the writing could be improved (e.g. “allows to : : :”, p2, line 18 and other instances, is not correct English). Proof-reading by a native English speaker would certainly help.

---

The text has been revised again, and any remaining grammar issues will hopefully be fixed by Copernicus’ language correction services at publication.

---

### 2 Detailed comments

---

Page 1, line 8: cross-validation against which data set? Please clarify here.

---

The sentence in the abstract has been updated as suggested by the reviewer.

**“The method’s validity was assessed through an extensive cross-validation against a dataset of 32 glaciers, with an estimated average error....”**

---

Page 1, line 16: I would not refer to “meltwater contributions” here: Annual values of glacier-wide SMB do not actually yield “meltwater” but just annual glacier mass loss. Meltwater has a strong seasonal component and also occurs in the case of SMB=0 or SMB >0!

---

Authors reply to Anonymous reviewer 1 on “A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015”

Indeed, the annual glacier-wide SMB is the net annual mass change of the glacier, so it is not precise to refer to it as meltwater contributions. The sentence has been updated accordingly.

“...provides relevant and timely data for studies in the fields of glaciology, hydrology and ecology in the French Alps, in need of regional or glacier-specific **annual net glacier mass changes** in glacierized catchments.”

---

- Page 2, line 12: order references according to year.

---

The references have been updated as suggested.

---

Page 2, line 13: Digital Elevation Model, instead of “maps”, is typically used

---

The acronym has been updated as suggested.

---

Page 2, line 22: This is also true for various other global glacier models (see Hock et al 2019 for a compilation). All of these models provide annual SMB for French glaciers in the PAST (although with probably limited skill). See also comment above. The detailed regional modelling study including the French Alps (Zekollari et al 2019) should also be mentioned.

---

Indeed, any global past SMB simulations include the European Alps with the French Alps, but these two specific studies were chosen since they were dedicated publications on the European Alps (Huss 2012, Marzeion et al., 2012) and they covered the full period from this study (1967-2015).

The study from Zekollari et al. (2019) was not used for comparison as the main purpose of their paper was to present the future evolution of the glaciers. The study covers the past period between 2003 until 2017, but this is a minimal fraction of time period of our study. It was not clear to us if annual glacier-wide SMB data is available for the 1967-2015 period, as the period seems to only be covered during validation, so it might only include glaciers with WGMS observations. Nonetheless, since all studies available use substantially different climatic forcings and SMB modelling approaches to our study, the type of comparison would still be the same, only serving to showcase the different sensitivities and responses of models to the past climate in the French Alps.

We have updated the sentence in order to give some context as suggested by the reviewer:

“On the other hand, SMB reconstructions have already been carried out in the European Alps, **providing a basis for comparison between different approaches (see Hock et al.**

Authors reply to Anonymous reviewer 1 on “A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015”

**(2019) for a compilation). Two studies include reconstructions in the European and thus the French Alps over a substantial period of the recent past: ...”**

---

Page 2, line 28: It seems strange to refer to the results of the present paper already within the paper itself. It is clear that the data set is already available, but it should not be referred to, as this is where it is actually described.

---

We understand that this might seem strange, and this would probably not be done in another journal. But the instructions for manuscript preparation of ESSD say:

“**Data sets:** The data sets described in the manuscript need to be deposited in [reliable data repositories](#) including the assignment of digital object identifiers. Authors are required to properly cite the data sets in the abstract, text, and the reference list (see section References below)”

Therefore, since here we are referring to the dataset itself, which is different from the paper (which acts only as a presentation), we believe that the citation makes sense according to the author guidelines. Nonetheless, if the reviewer and editor think this is not the correct way to use dataset citations for this journal we will remove it.

---

Figure 1: Although it is stated in the caption that the figure is schematic, it leaves a wrong impression on the density of available information for training the DL approach: In fact, green lines should only make up for 5% of the glaciers, whereas the figure implies that it is more than a third. It should be revised accordingly.

---

Indeed, the representation was not accurate. The number of glaciers with observations has now been reduced in Fig. 1. Nonetheless, it does not exactly account for 5% (as it would leave only one glacier which would not help to convey the message), but the representation is much more accurate now.

---

Page 3, line 10: Although this paper is closely related with TC paper of the same authors, the location of the training data sets needs to be presented here.

---

The methods section of this paper is a brief summary of the whole Bolibar et al. (2020) paper. We believe that the exact location of the 32 training glaciers is not very relevant for the reader to understand the methods used. The most important fact regarding the location is the fact that the training glaciers are distributed along most of the glacierized massifs of the French Alps, thus presenting a representative spatial coverage. As previously mentioned in other comments, for all the details regarding the methods the reader can refer to the dedicated methodological paper. Our intention is to avoid repeating unnecessary information, and to allow the reader to quickly understand the methods and the implications on the results without bothering too much with the details.

Authors reply to Anonymous reviewer 1 on “A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015”

Following the reviewer’s suggestion, and in order to make the paper fully self-sufficient, we have included the map of the study area with the glaciers used for training in the supplementary material.

“For the reconstruction presented here, a dataset of 32 French alpine glaciers has been used for training, covering most of the massifs within the French Alps, which exhibit a great variability of topographical characteristics (**Fig. S10**).”

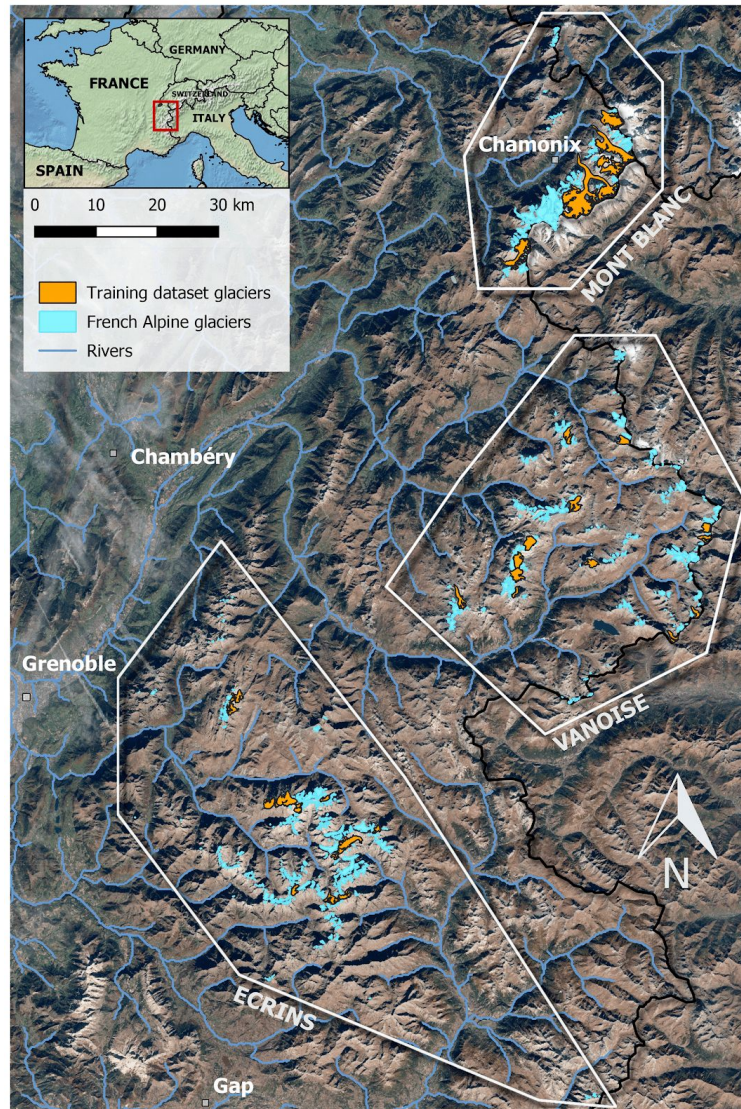


Figure S10. French Alpine glaciers used for model training and validation and their classification into three clusters or regions (Écrins, Vanoise, Mont-Blanc). Coordinates of bottom left map corner: 44°32' N, 5°40' E. Coordinates of the top right map corner: 46°08' N, 7°17' E.



---

Page 3, line 16: It is excellent that topographical information is available in repeated time steps throughout the study period. However, it remains unclear how this updated topographical data was included in the DL approach. This is quite relevant information as the feedback of retreating glaciers (shrinking area, changes in area-elevation distribution and terminus position) exerts an important effect on glacier-wide SMB. A detailed description of this is required.

---

We extended the comments in this section in order to make the presentation of the training predictors of the model clearer.

“Out of the 32 glaciers from this dataset, four glaciers include direct SMB measurements from the GLACIOCLIM observatory, some of which between 1949 and 2018 (Vincent et al., 2017) and 28 glaciers include estimates of annual glacier-wide SMB from remote sensing between 1984 and 2014 (Rabatel et al., 2016). **This dataset, with a total of 1048 annual glacier-wide SMB values, is used as a reference. Unlike point SMB, glacier-wide SMB is influenced by both climate and topography, producing complex interactions between climate and glacier morphology which need to be taken into account in the model. For each annual glacier-wide SMB value available, the following data are compiled to train the ANN with an annual time step: (1) climate data from the SAFRAN meteorological reanalyses (Durand et al., 2009) with: cumulative positive degree days (CPDD), cumulative winter snowfall, cumulative summer snowfall, mean monthly temperature and mean monthly snowfall, all variables being quantified at the altitude of the glacier's centroid; and (2) annually interpolated topographical data between the 1967, 1985, 2003 and 2015 glacier inventories in the French Alps (Gardent et al., 2014), with: mean and maximum glacier altitude, slope of the lowermost 20% altitudinal range of the glacier, surface area, latitude, longitude and aspect. Therefore, the topographical feedback of the shrinking glaciers is captured from these annually interpolated topographical predictors.** These parameters were identified as relevant for glacier-wide SMB modelling in the French Alps (Bolibar et al., 2020) and the dates of the glacier inventories determined the time interval for the reconstructions presented here.”

---

Page 4, line 1: Similar comment as above: I would just refer to the paper where the model is described and not have separate references to the model and the publication. You can state in the Data availability section where the code of the model is located.

---

Bolibar et al. (2020) is not a presentation of the model, but a presentation of the deep learning SMB modelling approach, with a case study on the French Alps. Therefore, when referring to the model itself (software), we prefer to use the citation for the model, as it is more precise. The Data availability section already includes both the dataset and the model's source code.

---

Page 4, line 15: It is a quite relevant aspect in my opinion (probably worth mentioning in the Intro) that the DL approach (e.g. in comparison to in-situ observations and mass balance modelling) only provides annual SMB, but no information on seasonal mass balance as well as mass balance gradients. If my understanding is wrong, please correct me. But these additional variables are crucial for various aspects of impact assessment and model development.

---

We agree that this aspect is important. This is mentioned in different sections of the manuscript, in the abstract, introduction and in some sections. We have tried to clarify this by systematically referring to the reconstructed SMB as “annual glacier-wide SMB”. Moreover, in the update from the reviewer’s comment on Page 3 Line 16, we have contrasted this fact with the characteristics of point mass balance data. If the reviewer thinks it would be clearer to refer to this in another way than “annual glacier-wide SMB” we could adapt the manuscript accordingly.

Seasonal mass balances are indeed very useful for several applications (see for instance Viani et al., 2018 in the field of hydrology). However in our case, the use of annual glacier-wide SMB data is not a problem, since ALPGM uses the delta-h (Huss et al., 2010) parameterization in order to redistribute the annual glacier-wide mass changes along the glacier. This geometry update is only used for glacier evolution simulations, which have nothing to do with the dataset presented here, so this information has been omitted in the methods from the present study.

---

Figure 2/5: Please add letters (A/B) to the panels.

---

Letters have been added to the different panels of Figures 2 and 5.

---

Page 7, line 18: For a more recent reference on exactly this topic, please also see Zekollari et al. (2020), *Geophysical Research Letters*

---

Indeed. A reference has been added to this paper.

---

Page 8, line 2: Please remove the reference to Huss&Hock, 2015. It does not fit here in my opinion.

---

Well spotted. This is in fact a LaTeX reference typo. The intended reference was: Huss et al. (2015): “New long-term mass-balance series for the Swiss Alps”, *Journal of Glaciology*. The citation has been updated accordingly.

---

Page 8, line 31: see also substantive comment above: To me, this appears too long and too detailed. The comparison to a modelling study is a nice addition but it does not actually allow evaluation of the present results. More focus needs to be put on the validation using fully independent field observations (in-situ) and geodetic surveys.

---

In order to make this section lighter and more straight to the point, we have moved a whole paragraph which compared and detailed the differences between both models to the Supplementary material. This allows to convey the message that a direct comparison is not possible, jumping straight to results and conclusions and leaving the technical details for the avid reader who will be willing to read the Supplementary material.

The new section S2 is the following one:

“2 Model differences between the updated version of Marzeion et al. (2015) and this study

In order to contrast the results from Sect. 3.4, three important different aspects between our approach and the one of M15U need to be highlighted:

1.  $M_{15U}$ 's model works with simplified physics, with a temperature-index model calibrated on observations; in this study we used a fully statistical approach based on deep learning, where physics-based considerations only appear in the predictor selection.
2.  $M_{15U}$  calibrated their model with SMB observations of 38 glaciers, most of them located in Switzerland for the 1901-2013 period; in this study we used observations of 32 glaciers, all located in the French Alps for the 1967-2015 period.
3.  $M_{15U}$  forced their updated model with CRU 6.0 (update of Harris et al., 2014), with  $0.5^\circ$  latitude/longitude grid cells, which has a significantly lower spatial resolution and suitability to mountain areas than the SAFRAN reanalysis (Durand et al., 2009) used in this study, in which altitude bands and aspects are considered for each massif, and meteorological observations from high-altitude stations are assimilated.

The cross-validations of both studies determined a performance with an average RMSE of  $0.66 \text{ m.w.e. a}^{-1}$  and an  $r^2$  of 0.43 for  $M_{15U}$  for the European Alps, and an average RMSE of  $0.49 \text{ m.w.e. a}^{-1}$  and an  $r^2$  of 0.79 for this study. However, due to the highly different methodologies and forcings of the two models, a direct comparison is not possible, so the following analysis is focused on the overall trends and sensitivities in the reconstructions and their potential sources.”

---

Page 10, line 3: Potentially also related to the way glacier retreat (updating of area elevation distribution) is accounted for in the model by Marzeion et al 2015, and the present approach? Probably worth discussing here as well.

---

That's a very good point. On top of the differences in the climate forcings, there might be differences in the topographical feedback of the models due to the different modelling approaches. This element has been added to the discussion:

Authors reply to Anonymous reviewer 1 on “A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015”

“The fact that  $M_{15U}$  used a volume-area scaling compared to the interpolated topographical data from inventories from this study means that the topographical feedback of the models might differ as well throughout the reconstructed period.”

---

Page 11, line 6: Actually, all Supplementary Figures should be referenced from the main text. I found the analysis in the Supplementary interesting but not straightforward to understand. It might be beneficial to present this additional analysis more prominently in the main text.

---

All Supplementary Figures that were not previously mentioned in the main text have now been added as references in the appropriate sections. With this we hope the reader will be encouraged to read the supplementary material in case she/he is interested in the detailed methods. By giving the references in the right context, it is now easier to relate the explanations of Sect. 2 in the Supplementary to the content of the main text. Our intention is to keep the methods section and technical details as light as possible, in order to convey an easy message based on the results, and allow the avid reader to check the details in the methods paper and the Supplementary material.

---

Page 12, line 12: Is there no possibility to go beyond the year 2015 by the way? In my understanding the trained DL approach should enable to predict mass balance also for the most recent years. This would be quite interesting as the last years were extraordinary in terms of their mass losses.

---

Indeed, it would be very interesting, but unfortunately with the current framework it cannot be done, unless some modifications are done to it. Since we are working with topographical data by interpolating glacier inventories, we would need a glacier inventory after 2015 in order to have topographical information for these years. One possible hypothesis to bypass this would be to continue interpolating the topographical data with the same trend as for the 2003-2015 period, but this would introduce some inhomogeneity in the method. On the other hand, the version of the meteorological reanalysis that we are using (SAFRAN) only extends until 2016. A new version has just been released, that covers the same period until 2019 ; its use would require some technical adaptations and reprocessing of the years previous 2015 for homogeneity. Considering these two facts, we believe it is not worth the time investment for just four additional years.