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

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 ± 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 21st 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.
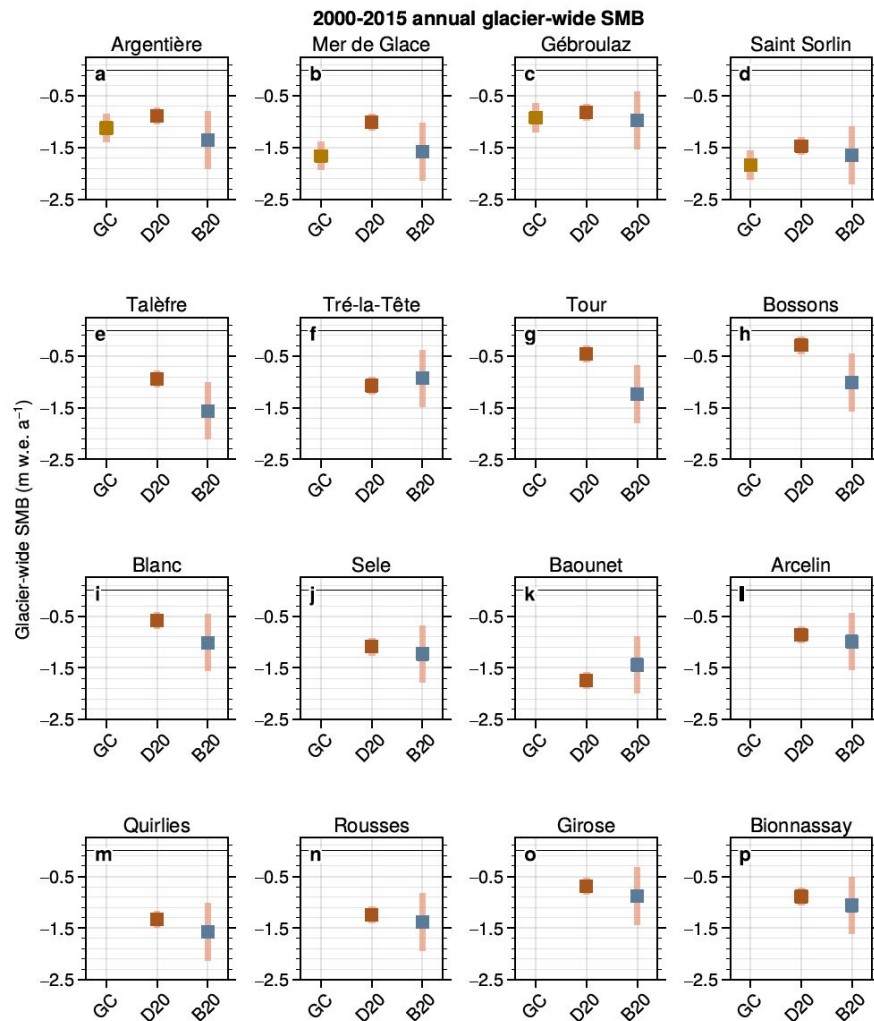
A new section analyzing this ("1 Comparison with independent geodetic mass balance data"), including the two following figures have been added to the supplementary material, in order to illustrate this.

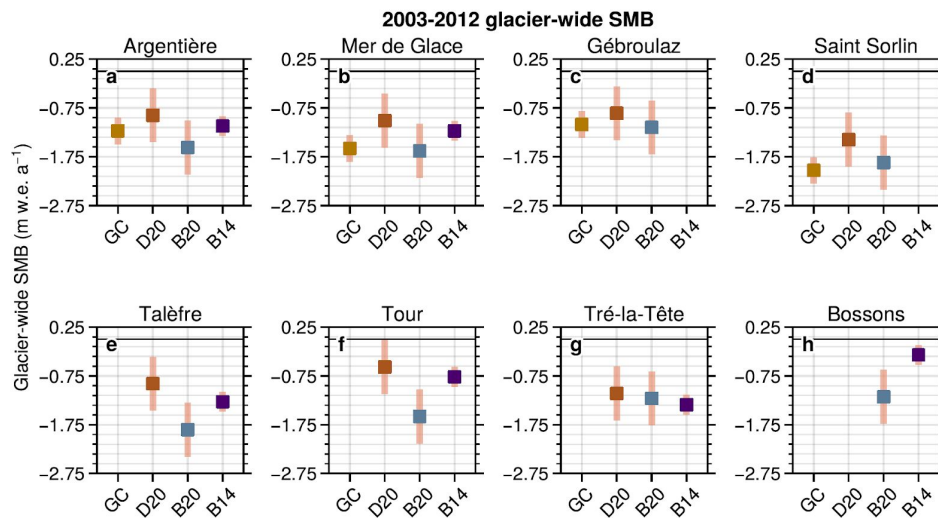"1 Comparison with independent geodetic mass balance data

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

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

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."

2000-2015 annual glacier-wide SMB

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!

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.**

(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)**."
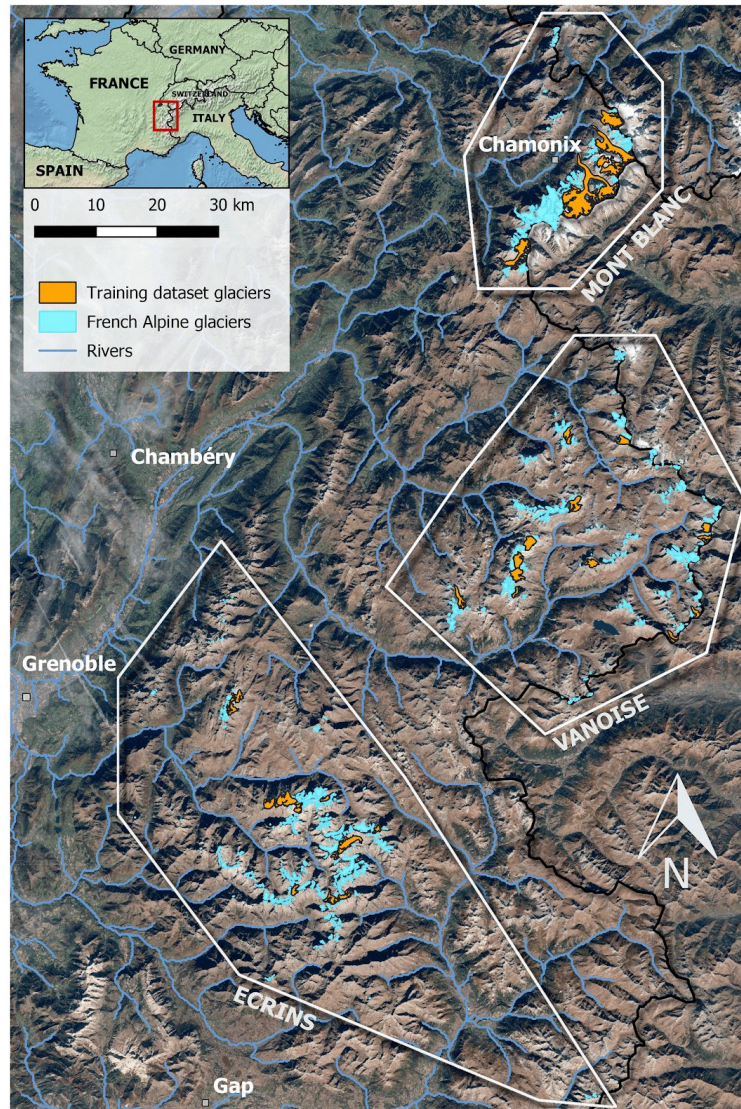
Figure S10. French Alpine glaciers used for model training and validation and their classification into three clusters or regions (Écrins,Vanoise, Mont-Blanc). Coordinates of bottom left map corner: 44º32' N, 5º40' E. Coordinates of the top right map corner: 46º08' N, 7º17'E.

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.

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° 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 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 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."

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.

Authors reply to Ben Marzeion on "A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015"

Ben Marzeion

1 General comments

Bolibar et al. present the results of a new approach to reconstruct glacier mass balances at times and/or locations where meteorological conditions (and some topographical information) are known, but no observations of glacier mass balance exist. Their approach, based on a neural network algorithm, adds considerable diversity to the existing group of reconstruction methods. The thorough validation of the results leads to great confidence in the robustness of the method. Except for some minor issue listed below, the manuscript is very clear and easy to follow. The data set produced and presented here will be of great use for the community. I particularly appreciate the great care that has been taken in documenting the test for overfitting in the supplementary material. I recommend publication once the authors have gone through the list of questions/suggestions below.

We are grateful for the positive and encouraging comments. These comments will help improve the manuscript's quality and clarity. Most figures in the paper have been re-processed taking into account the feedback, hopefully leading to better visualization and presentation. Every comment/suggestion has been addressed individually in the following section.

As explained in one of the comments regarding the validation approach based on cross-validation, 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.

2 Specific/minor comments

P1 L9: please specify "1\sigma" instead of "\sigma" for clarity.

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

P1 L10: the "moderately" should only apply to the 1980s, I think

Authors reply to Ben Marzeion on "A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015"

Indeed. The sentence has been updated:

"We estimate an average regional area-weighted glacier-wide SMB of -0.72±0.20 (1\sigma) m.w.e. a^{-1} for the 1967-2015 period, **with negative mass balances** in the 1970s (-0.52 m.w.e. a^{-1}), **moderately negative** in the 1980s (-0.12 m.w.e. a^{-1}), and an increasing negative trend from the 1990s onwards, up to -1.39 m.w.e. a^{-1} in the 2010s."

P1 L10: avoid line break within negative number

This has been fixed with the rephrasing of some parts of the abstract.

P1 L12: unclear, what "this period" refers to

The sentence has been updated to clearly indicate the time period:

"Following a topographical and regional analysis, we estimate that the massifs with the highest mass losses **for the 1967-2015** period are the..."

---

abstract: why are no uncertainties given for the values of the different massifs? (also concerns the conclusions)

Because we have no way to dissociate the uncertainties for each massif from the overall uncertainties computed through cross-validation. Therefore, all massifs would display the same uncertainty, which is already given with the average performance of the method for this region (RMSE = 0.55 m w.e. a^{-1}). If the reviewer thinks it would still be better to give the uncertainty for each massif we can add it in the abstract and text.

P2 L8: "these points" refers to the points of MB measurements, but this reference is not very clear here; also, it's not the points that show nonlinear variability, but the measurements at the points; suggest to rephrase

This sentence has been adapted to improve clarity as suggested:

"**These different point SMB measurements** can show a high nonlinear variability..."

Authors reply to Ben Marzeion on "A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015"

P2 L23: there more four global parameters in the Marzeion et al. (2012) model, and I wouldn't necessarily say they were "optimized", because that "optimization" was very subjective...

The word "optimized" has been removed to erase these connotations from the sentence as suggested by the reviewer:

"They used a minimal model relying only on temperature and precipitation data, based on a temperature-index method, with two parameters to calibrate the temperature sensitivity and the precipitation lapse rate."

Fig. 1: the figure certainly works well for presentations etc., but I'm not sure it is necessary here, since the text describes very well what is done, and there is little to be gained from the figure.

Indeed, the main key aspects of the overall analysis are already given in the abstract and in the text. Nonetheless, we believe it is a complementary way to show the regional variability, as it shows in a single figure the spread of glacier behaviour and the common variability in a nice and easy way. This is our personal opinion, if the reviewer strongly suggests to remove it, we will move it to the supplementary material.

P3 L14-15: it would be great if you can add a sentence or two here, specifying how any difference in the altitude of the glaciers' centroids and the reanalysis grid points were treated (lapse rates or similar?)

The explanation on climate data and predictors has been updated with the following sentences in order to give some context on how the forcings are adjusted to each glacier centroid's altitude:

"(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. **In order to capture the climate signal at each glacier's centroid, temperatures are taken from the nearest SAFRAN 300 m altitudinal band and adjusted with a 6ºC/km lapse rate. The updated temperature is then used to update the snowfall amount from the same 300 m altitudinal band.**"

P4 L22 or lower: It might be worth pointing out/discussing that the density of observations used in the LOGO cross validation is denser towards the end of the reconstruction interval, when presumably, also the quality of the meteorological data are higher, such that the uncertainty of the methods might be underestimated for the (roughly) first half of the period. I also wonder if/how this interferes with your assessment of the model's ability to reconstruct the more neutral MB values during 1967-1984?

That's a very good point. That was one of our main concerns during the validation process, which we tried to address in two different ways.

First of all, we performed a separate cross-validation with only data from the 1967-1984 period, in order to specifically assess the performance during this period. This is explained in the newly created Sect. "2.3 Uncertainty assessment" (as suggested by Anonymous reviewer 1). This was already present in the version of the manuscript sent for review.

On the other hand, in order to improve our estimates and to better take into account this lack of homogeneity in the dataset, we have trained a new ensemble of models based on Leave-Some-Years-and-Glaciers-Out (LSYGO) cross-validation, as explained in Bolibar et al. (2020). We used an ensemble of 60 CV models using weighted bagging (Hastie et al., 2009) by giving +33% more weight to data between 1967-1984, in order to compensate for this lack of observations during this period, which covers a third of the 49-year period. This has not affected much the results, and the conclusions remain exactly the same, but it allows giving a more accurate and realistic assessment of the model's performance, with a RMSE of 0.55 m w.e. a^{-1}, a coefficient of determination of 0.75 and an average bias of -0.019 m w.e. a^{-1}.

Fig. 2: since there are so many lines, it is somewhat hard to see the distribution. Particularly in the lower panel, a histogram for showing the distribution of the accumulated values (vertically, to the right of the panel) would be quite interesting. It would be possible to see, e.g., how/if the area weighted mean differs from the "ensemble" mean and/or median, if the distribution is (a)symmetric, etc. Just a suggestion to consider.

That is a good idea. A panel to the right of the cumulative plot has been added with a histogram, the PDF and the position of the area weighted mean SMB.

Authors reply to Ben Marzeion on "A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015"
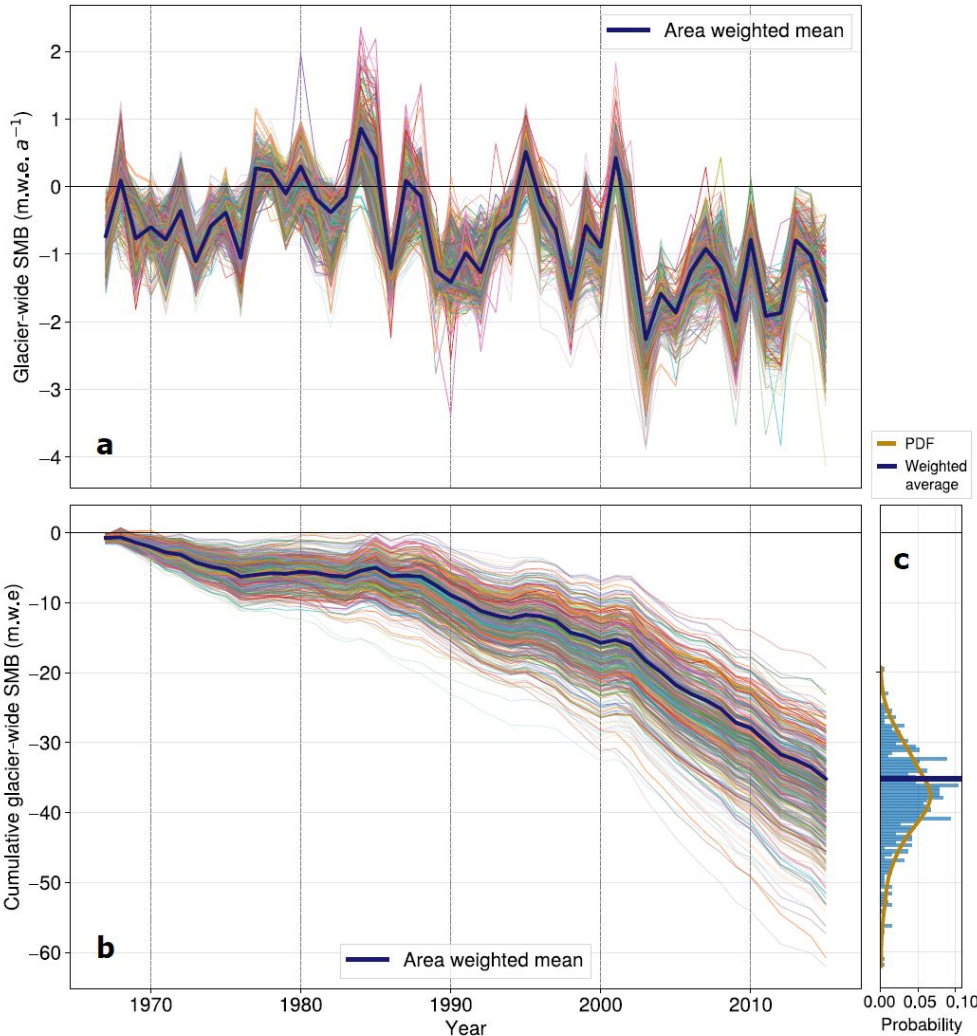

Fig. 3: why are no uncertainties included for the decadal averages?

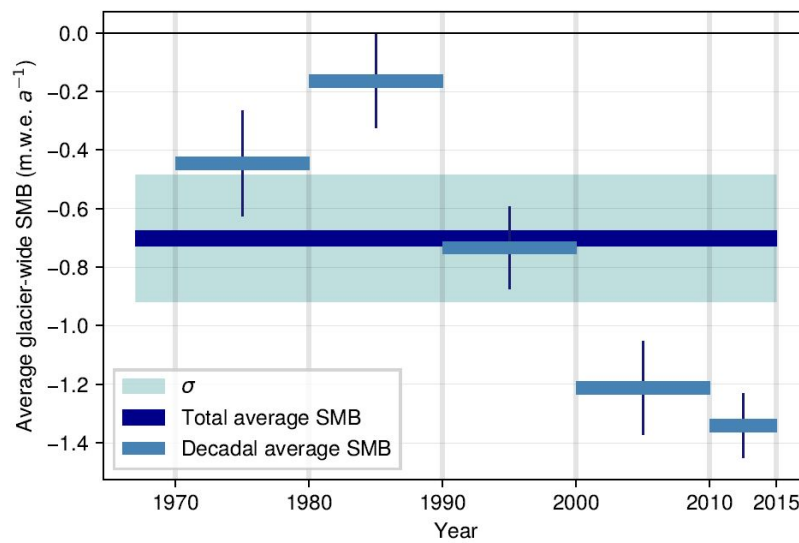Fig. 3 has been updated with decadal uncertainties.

Authors reply to Ben Marzeion on "A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015"

Fig. 4: great figure! But a bit busy (just visually); would it be possible to mute the background image a bit (and then perhaps change the text color to black) so that the colors of the glaciers stand out more?

We do agree that Fig. 4 could be a little bit overwhelming. We have updated it following the suggestions of the reviewer. Now the colours of the glaciers are more visible, and it has a more homogenous feeling with all contour lines in black.

Authors reply to Ben Marzeion on "A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015"
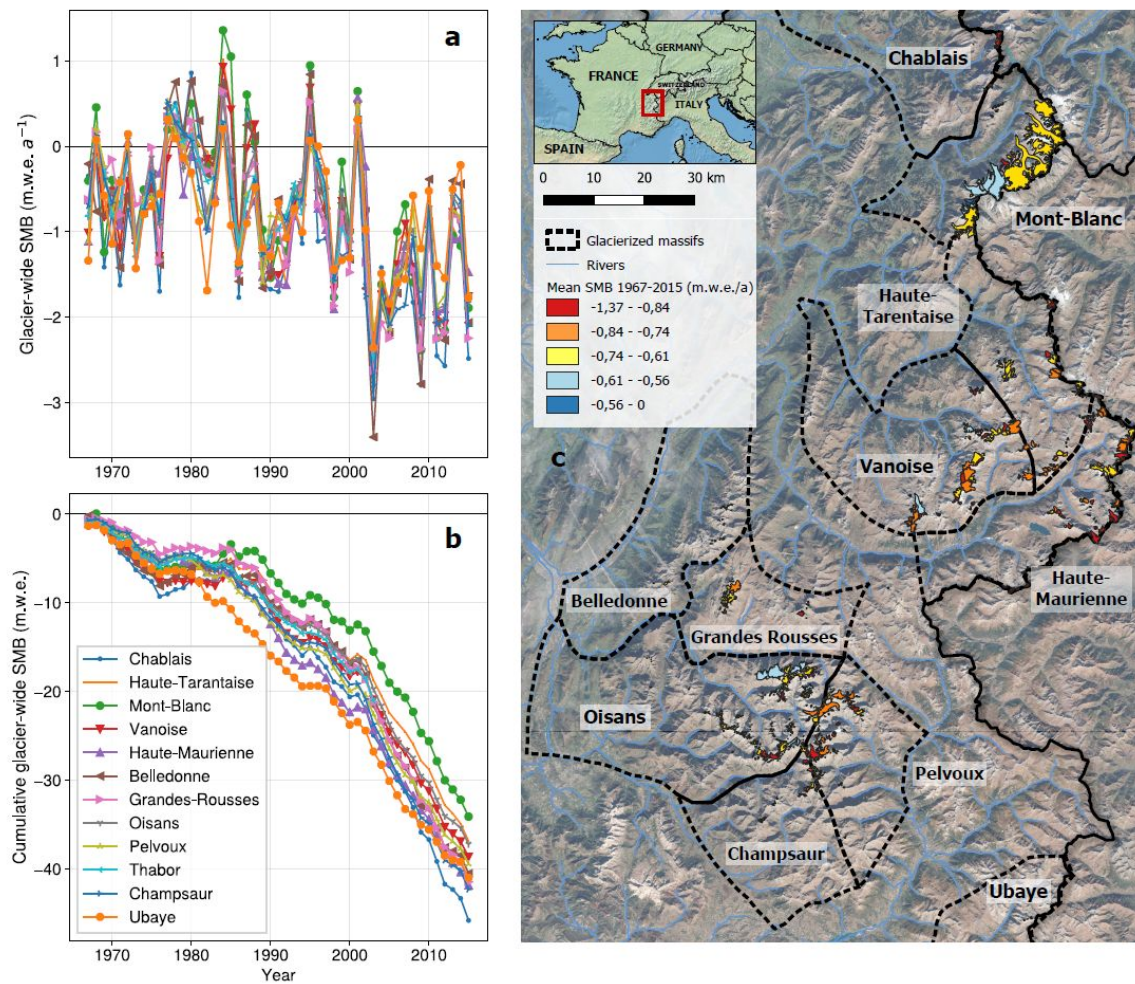
P8 L16: it's more than three parameters: one local (the temperature sensitivity) and four global ones (precipitation correction factor, precipitation lapse rate, temperature threshold for solid precipitation, and melt temperature threshold); see Figs. 4-7 in Marzeion et al. (2012).

The sentence has been updated with the correct information as it follows:

"This model was optimized based on five parameters: the temperature sensitivity of the glacier (local); and a precipitation correction factor, precipitation lapse rate, temperature threshold for solid precipitation and melt temperature threshold (global)"

P8 L22: perhaps clarify that the 38 glaciers are not the global sample used for calibration.

The new section in the supplementary material has been updated following this suggestion:

Authors reply to Ben Marzeion on "A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015"

"**M15U calibrated their model with global SMB observations, including 38 glaciers in the European Alps, 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."

P8 L31: I believe that the CV results in the Marzeion et al. (2012) study are also influenced by the global "optimization" (see above) of the four parameters; probably, a focus on the Alps would have led to a different parameter choice, and hence different CV results.

Indeed, that is what we tried to convey with the warning to the readers. We are comparing a global model with a regional model, so the specificity of the calibration is completely different, giving a clear advantage to the regional model. We hope that with the updated sentence from the previous comment this will be more clear to the reader.

P10 L1 and following: another reason for the different behaviour around the 2003 "break point" might be that the Marzeion et al. (2012) model, by construction, cannot capture the lasting effect that the extreme 2003 year may have had on albedo; while your model may be able to capture this (I guess – I'm not sure) by essentially taking the time as an additional predictor?

Our mass balance model does not have any perception of time, as no time stamps are used as predictors. I believe the main reason(s), as stated in the article, are the fact that we use higher resolution climate forcings, which better capture the climate signal on the glaciers, and most importantly, that the deep learning SMB model is nonlinear, which gives it a greater deal of flexibility to simulate this kind of transitions compared to linear models. This was already observed during the cross-validation analysis in Bolibar et al. (2020), where the linear model with Lasso, which behaves similarly to a temperature-index model, showed biases at the beginning and end of the 1984-2015 period, as the parameters were calibrated to fit the whole period, which presents rather neutral SMBs at the beginning and strongly negative SMBs by the end (see the Figure below taken from Bolibar et al., 2020). The nonlinear deep learning SMB model showed much lower biases, further demonstrating that the climate and glacier systems are highly nonlinear.

Authors reply to Ben Marzeion on "A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015"
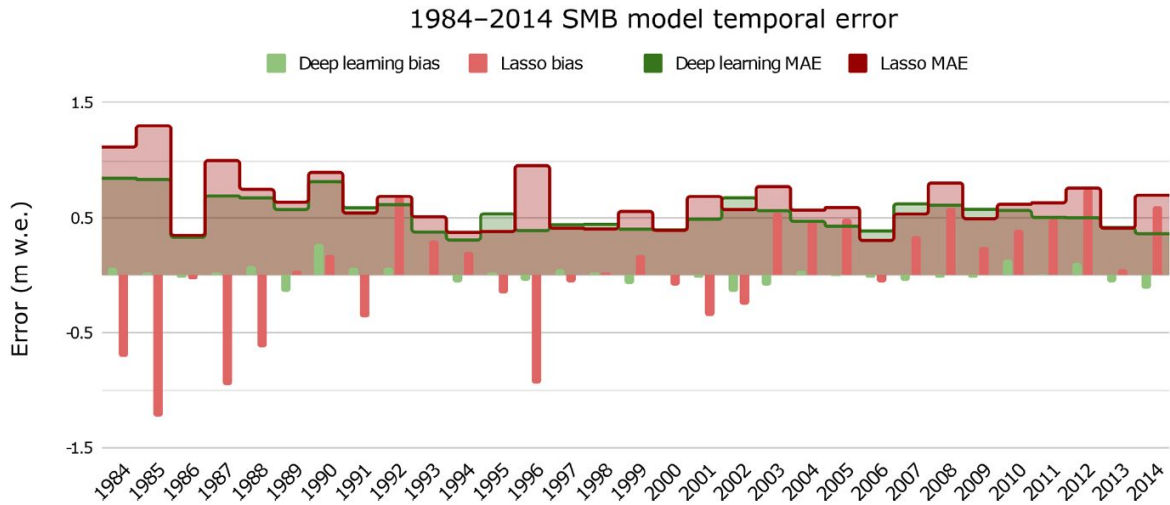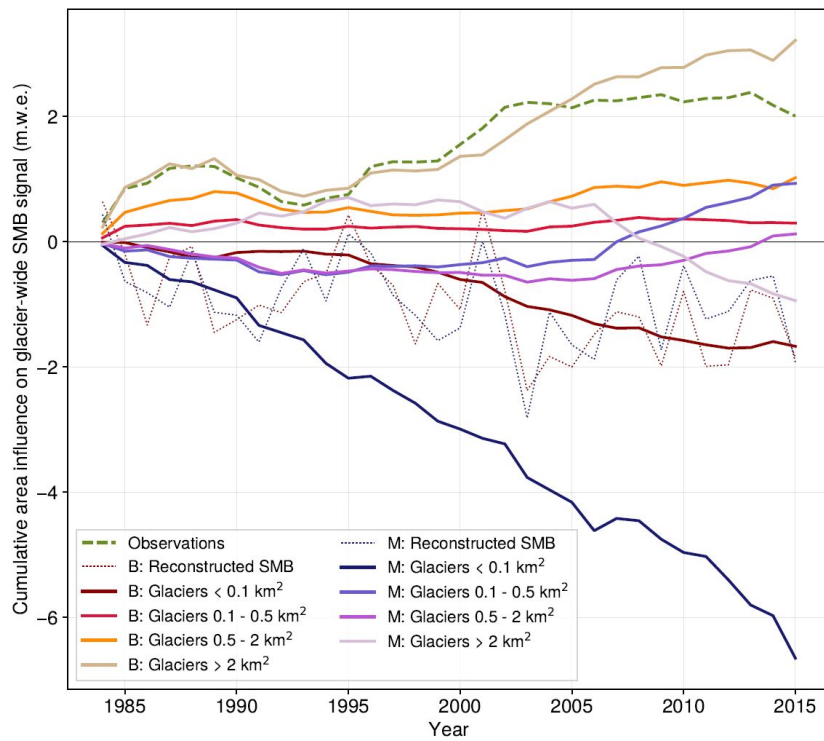

1984–2014 SMB model temporal error

Fig. S2: would it be possible to re-arrange the legend such that it is easier to compare the "B" to the "M" lines (e.g., shift the lowest line in the legend to the right)?

The legend in Fig. S4 (previously S2) has been updated in order to have the "B" and "M" lines on different columns.

A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967-2015

Jordi Bolibar[1,2], Antoine Rabatel[1], Isabelle Gouttevin[3], and Clovis Galiez[4]

[1]Univ. Grenoble Alpes, CNRS, IRD, G-INP, Institut des Géosciences de l'Environnement (IGE, UMR 5001), Grenoble, France
[2]INRAE, UR RiverLy, Lyon-Villeurbanne, France
[3]Univ. Grenoble Alpes, Université de Toulouse, Météo-France, CNRS, CNRM, Centre d'Études de la Neige, Grenoble, France
[4]Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, Grenoble, France

Correspondence: Jordi Bolibar (jordi.bolibar@univ-grenoble-alpes.fr)

Abstract. Glacier surface mass balance (SMB) data are crucial to understand and quantify the regional effects of climate on glaciers and the high-mountain water cycle, yet observations cover only a small fraction of glaciers in the world. We present a dataset of annual glacier-wide surface mass balance of all the glaciers in the French Alps for the 1967-2015 period. This dataset has been reconstructed using deep learning (i.e. a deep artificial neural network), based on direct and remote sensing SMB observations, meteorological reanalyses and topographical data from glacier inventories. This data science reconstruction approach is embedded as a SMB component of the open-source ALpine Parameterized Glacier Model (ALPGM). ~~An extensive~~ The method's validity was assessed through an extensive cross-validation ~~allowed to assess the method's validity~~ against a dataset of 32 glaciers , with an estimated average error (RMSE) of ~~0.49~~ 0.55 m.w.e. a^{-1}, an explained variance (r^2) of ~~79~~75% and an average bias of ~~+0.017~~ -0.021 m.w.e. a^{-1}. We estimate an average regional area-weighted glacier-wide SMB of ~~-0.72~~ -0.71±~~0.20~~ 0.21 (1σ) m.w.e. a^{-1} for the 1967-2015 period, with ~~moderately~~ negative mass balances in the 1970s (~~-0.52~~ -0.44 m.w.e. a^{-1})~~and~~ , moderately negative in the 1980s (~~-0.12~~ -0.16 m.w.e. a^{-1}), and an increasing negative trend from the 1990s onwards, up to ~~-1.39~~ -1.34 m.w.e. a^{-1} in the 2010s. Following a topographical and regional analysis, we estimate that the massifs with the highest mass losses for ~~this~~ the 1967-2015 period are the Chablais (~~-0.90~~ -0.93 m.w.e. a^{-1})~~and Ubaye and Champsaur ranges (-0.91~~ , Champsaur and Haute-Maurienne (-0.86 m.w.e. a^{-1} both) and Ubaye ranges (-0.83 m.w.e. a^{-1}~~both~~), and the ones presenting the lowest mass losses are the Mont-Blanc (~~-0.74~~ -0.69 m.w.e. a^{-1}), Oisans and Haute-Tarentaise ranges (~~-0.78~~ -0.75 m.w.e. a^{-1} both). This dataset - available at: https://doi.org/10.5281/zenodo.3663630 (Bolibar et al., 2020a) - 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 ~~meltwater contributions~~ annual net glacier mass changes in glacierized catchments.

1 Introduction

Among all the components of the Earth system, glaciers are some of the most visibly affected by climate change, with an overall worldwide shrinkage despite important differences between regions (Zemp et al., 2019). The European Alps are among the

regions with the strongest glacier mass loss over recent decades, with expected mass losses between 60% and 95% by the end of the 21st century (Zekollari et al., 2019). These major glacier mass changes are likely to have an impact on water resources, society and alpine ecosystems (e.g. Huss and Hock, 2018; Immerzeel et al., 2020; Cauvy-Fraunié and Dangles, 2019). In order to study and quantify all these potential consequences, the availability of glacier mass balance data is of high relevance. Therefore,

5 open historical datasets are crucial for the understanding of the driving processes and the calibration of models used for projections. Unlike glacier length, glacier surface mass balance (SMB) provides a more direct indicator of the climate-glacier interactions (Marzeion et al., 2012). Glacier SMB is classically measured using the direct or glaciological method, by separately determining the ablation and accumulation totals. Direct measurements quantify the surface mass balance at different points of the glacier, and these values must be integrated at the glacier scale in order to assess the glacier-wide SMB (Benn and Evans, 2014).

10 These different ~~points~~ point SMB measurements can show a high nonlinear variability, which can complicate this integration process towards glacier-wide estimates (Vincent et al., 2018). Moreover, field measurements require a lot of manpower, time and economic resources in order to be sustained for a meaningful period of time. On the other hand, recent advances in remote sensing allow estimating glacier SMB changes at a regional level with unprecedented efficiency using geodetic and gravimetric methods ~~(Kääb et al., 2012; Berthier et al., 2016; Fischer et al., 2015; Brun et al., 2017; Dussaillant et al., 2019)~~(Kääb et al., 2012; Fischer

15 . Due to constraints related to the availability of digital elevation ~~maps~~ models (DEMs) or airborne data, these mass balance estimates normally encompass several years or decades. Some studies are bridging the gap towards an annual temporal resolution (Rabatel et al., 2005, 2016; Rastner et al., 2019), but the coverage is still limited to glaciers without cloud cover or acquisition-related artefacts. This means that these mass balance datasets are often restricted to certain glaciers and years within a region. All these new datasets are extremely beneficial for data-driven approaches, fostering the training of machine learning models

20 capable of capturing the regional characteristics and relationships (Bolibar et al., 2020b). This type of approach allows to fill the spatiotemporal gaps in the SMB datasets, therefore, it can be seen as a complement to remote sensing and direct observations.

On the other hand, SMB reconstructions have already been carried out in the European Alps, providing a basis for comparison between different approaches ~~. For example,~~ (see Hock et al. (2019) for a compilation). Two studies include reconstructions in the European and thus the French Alps over a substantial period of the recent past: Marzeion et al. (2012, 2015) recon-

25 structed annual SMB series of all glaciers in the Randolph Glacier Inventory ~~- including the European Alps -~~ for the last century. They used a minimal model relying only on temperature and precipitation data, based on a temperature-index method, with two ~~optimized~~ parameters to calibrate the temperature sensitivity and the precipitation lapse rate. Huss (2012) presented an approach to extrapolate SMB series of a limited number of glaciers to the mountain-range scale. By comparing multiple methods, he found the best results with a multiple linear regression based on 6 topographical parameters. From this relationship

30 he reconstructed area-averaged SMB series of all the glaciers of the European Alps between 1900-2100 and analysed the trends for the different alpine nations and different glacier sizes.

Here, we introduce a dataset of annual glacier-wide SMB of all the glaciers in the French Alps (Bolibar et al., 2020a), located in the westernmost part of the European Alps, between 5.08° and 7.67°E, and 44° and 46°13'N. Glacier-wide SMBs have been reconstructed for the 1967-2015 period, using deep learning (i.e. a deep artificial neural network). This approach was

35 introduced in Bolibar et al. (2020b), for which a deep artificial neural network (ANN) was trained with data from 32 French
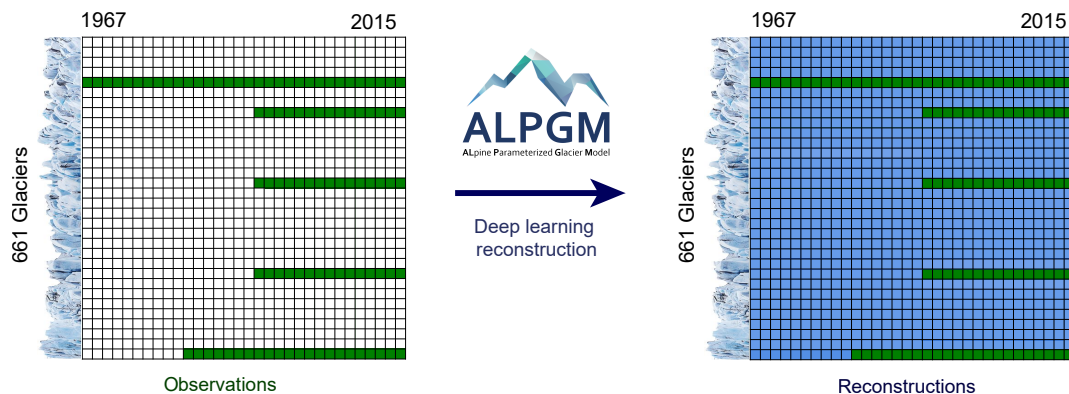
Figure 1. Summary of the deep learning regional SMB reconstruction approach. From the available annual glacier-wide SMB observations, a deep learning model is used to reconstruct the full dataset, thus filling the spatiotemporal gaps in the observational dataset. Green indicates glaciers and years with SMB observations and blue indicates reconstructed SMB values. The grid size with glaciers and years is schematic and only serves to illustrate the concept.

alpine glaciers, as part of the ALPGM glacier evolution model (Bolibar, 2020). Annual glacier-wide SMB values are reported for each glacier in the French Alps found in the 2003 glacier inventory (Gardent et al., 2014). An overview of the methodology used to produce the dataset and a review of the associated uncertainties is presented in Sect. 2, followed by a dataset overview in Sect. 3, where the data structure and regional trends are described and where the dataset is compared to a previous study and observations.

2 ~~Methodology~~ Data and ~~uncertainties~~methods

~~The annual glacier-wide SMB dataset for the 661 French alpine glaciers has been reconstructed using a deep artificial neural network (ANN), also known as deep learning. ANNs are nonlinear statistical models inspired by biological neural networks (Fausett, 1994; Hastie et al., 2009). Recent developments in the field of machine learning and optimization enabled the use of deeper ANN architectures, which allow to capture more nonlinear and complex patterns in data even for small datasets (Ingrassia and Morlini, 2005).~~

2.1 Data

For the reconstruction presented here, a ~~deep feed-forward ANN has been trained with 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. 4 in Bolibar et al., 2020b). Four glaciers~~ (Fig. S10). The French Alps are located in the westernmost part of the European Alps, rising from the Mediterranean sea northwards between 44 and 46º13' N, 5.08 and 7.67º E. Due to

3

its particular geographical setup, glacierized mountain ranges in the French Alps have distinct climatic signatures. Southern glaciers exhibit a Mediterranean influence, whereas northern glaciers are mostly affected by western fluxes from the Atlantic, except for eastern glaciers close to the Italian border, which are more influenced by east returns.

Out of the 32 glaciers from this dataset, four glaciers include direct SMB measurements from the GLACIOCLIM observatory, some of which ~~between~~ since 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). ~~Training data consists of : (1)~~ This dataset, with a total of 1048 annual glacier-wide SMB ~~for each of the 32 glaciers~~ values, is used as a reference~~dataset; (2~~. 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 (3~~'s centroid. In order to capture the climate signal at each glacier's centroid, temperatures are taken from the nearest SAFRAN 300 m altitudinal band and adjusted with a 6 ºC/km lapse rate. The updated temperature is then used to update the snowfall amount from the same 300 m altitudinal band. (2) annually interpolated topographical data ~~from between~~ the 1967, 1985, 2003 and 2015 glacier inventories in the French Alps (update of 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. ~~These~~ Therefore, the topographical feedback of the shrinking glaciers is captured from these annually interpolated topographical predictors. . These topoclimatic parameters were identified as relevant for glacier-wide SMB modelling in the French Alps (Bolibar et al., 2020b), and the dates of the glacier inventories determined the time interval for the reconstructions presented here.

For more details on the choice of predictors, the reader can find a thorough analysis in Bolibar et al. (2020b).

2.2 Methods

The annual glacier-wide SMB dataset for the 661 French alpine glaciers has been reconstructed using a deep artificial neural network (ANN), also known as deep learning. ANNs are nonlinear statistical models inspired by biological neural networks (Fausett, 1994; Hastie et al., 2009). Recent developments in the field of machine learning and optimization enabled the use of deeper ANN architectures, which allows capturing more nonlinear and complex patterns in data even for small datasets (Ingrassia and Morlini, 2005). This modelling approach is part of the SMB component of ALPGM (Bolibar, 2020), an open-source data-driven parameterized glacier evolution model. For a detailed explanation of the methodology, please refer to Bolibar et al. (2020b). For the final reconstructions presented here, a cross-validation ensemble approach was used, in which the individual predictions of each of the ~~Leave-One-Glacier-Out (LOGO~~60 Leave-Some-Years-and-Glaciers-Out (LSYGO) cross-validation ~~models~~ model members were averaged to produce a single output. An ensemble approach has the advantage of further improving generalization, and reducing overfitting as well as the inter-model high variance typical from neural networks (Krogh and Vedelsby, 1995). A weighted bagging approach (Hastie et al., 2009) was used in order to balance the dataset, giving

4

more weight to under-represented data samples from the years 1967-1983. On the other hand, for the 32 glaciers with glacier-wide SMB observations used for training, an ensemble of 50 models trained with the full dataset was used, in order to achieve the best possible performance for this subset of glaciers, which represents a substantial fraction (45% in 2003) of the total glacierized surface area in the French Alps.

2.3 Uncertainty assessment

The uncertainties linked to the deep learning approach used in this study have been assessed through cross-validation, for which deep learning predictions were compared with observations and remote sensing estimates. A detailed presentation of the method's uncertainties and performance from the cross-validation study can be found in Bolibar et al. (2020b). Block cross-validation ensured that all the 32 glaciers in the dataset were evaluated, with spatiotemporal structures formed by glaciers and years being considered in order to prevent the violation of the assumption of independence (Roberts et al., 2017). This means that three different deep ANNs were produced: one for reconstructing glacier-wide SMB in space, one for the reconstruction in time (future and past), and another one for both dimensions at the same time; each of these with a different calibration and performance. It was shown that the deep ANN performs better in the spatial dimension, in which the SMB signal relationships with the predictors are the simplest. SMB interannual variability is mostly driven by climate, whereas geography and local topography (i.e. differences between glaciers) modulate the signal in space in a simpler way (Vincent et al., 2017; Bolibar et al., 2020b). Therefore, deep learning is capable of finding more structures in the spatial dimension, accounting for a better accuracy and explained variance compared to the temporal dimension. The deep ANN used in this study presents an RMSE of 0.49 0.55 m.w.e a^{-1} with an r^2 of 0.79 in LOGO 0.75 in LSYGO cross validation. Nonetheless, only one glacier in the training dataset is smaller than 0.5 km^2 (Glacier de Sarennes, 0.3 km^2 in 2003), implying that uncertainties for very small glaciers (< 0.5 km^2) might differ from those estimated using cross-validation. In 2015, very small glaciers in the French Alps represented about 80% of the total glacier number, but they accounted for only 20% of the total glacierized area. This means that their importance is relative, for example in terms of water resources, but a user of this dataset should bear in mind that SMB from these very small glaciers might carry greater uncertainties than the ones assessed during cross-validation. This might be especially true for extremely small glaciers (< 0.05 km^2) which can be considered as spatial outliers for the deep ANN. Since there is only one glacier with SMB observations for very small glaciers and none for extremely small glaciers, there is no precise way to quantify these uncertainties. On the other hand, the ANN is mostly trained with glacier-wide SMB data between 1984 and 2014, with a reduced amount of values between 1967 and 1984 (986 and 62 values, respectively). Since this early period contains on average more positive and neutral glacier-wide SMB values than the 1984-2014 period, the performance of the ANN was specifically assessed for this period. An additional cross-validation was performed with four folds, each with a glacier including glacier-wide SMB data before 1984. For each fold, all SMB data of that glacier and time period were hidden from the ANN, and the simulated glacier-wide SMBs between 1967 and 1983 were tested in order to assess the model's performance. The results showed that the ANN is capable of correctly reconstructing glacier-wide SMB for glaciers and years before 1984 (Fig. S3 S5), with an estimated accuracy (RMSE) of 0.47 m.w.e. a^{-1} and an estimated explained variance (r^2) of 0.65. This uncertainty assessment is based on roughly 10% of the full dataset, meaning that these estimates

lack the robustness of the full cross-validation from Bolibar et al. (2020b), but they serve to show that the model can accurately reconstruct glacier-wide SMB data outside the main cluster of years used during training.

In order to further validate the reconstructions presented here, a comparison against independent ASTER (Davaze et al., 2020) and Pléiades (Berthier et al., 2014) geodetic mass balance data has been performed to validate the bias of the SMB reconstructions for the 2000-2015 (Fig. S1) and 2003-2012 (Fig. S2) sub-periods. Our reconstructions show a good agreement with the geodetic mass balances, except for some quite particular high altitude glaciers from the Mont-Blanc massif that substantially differ from most glaciers in the French Alps. A more detailed analysis and the figures comparing the mass balance datasets can be found in the supplementary material.

3   Dataset overview

3.1   Dataset format and content

The SMB dataset is comprised of multiple CSV files, one for each of the 661 glaciers from the 2003 glacier inventory (Gardent et al., 2014), named with its GLIMS ID and RGI ID with the following format: *GLIMS-ID_RGI-ID_SMB.csv*. Both indexes are used since some glaciers that split into multiple sub-glaciers do not have an RGI ID. Split glaciers have the GLIMS ID of their "parent" glacier and an RGI ID equal to 0. Every file contains one column for the year number between 1967 and 2015 and another column for the annual glacier-wide SMB time series. Glaciers with remote sensing-derived observations (Rabatel et al., 2016) include this information as an additional column. This allows the user to choose the source of data, with remote sensing data having lower uncertainties (0.35 ± 0.06 (σ) m.w.e. a^{-1} as estimated in Rabatel et al. (2016)). Columns are separated by semicolon (;). All topographical data for the 661 glaciers can be found in the updated version of the 2003 glacier inventory included in the Supplementary material and in the dataset repository.

3.2   Overall trends

We estimate an average area-weighted regional glacier-wide SMB of ~~-0.72~~-0.71±~~0.20~~0.21 (σ) m.w.e. a^{-1} between 1967 and 2015 (Fig. 3). As reported in previous studies (Huss, 2012; Rabatel et al., 2016; Vincent et al., 2017), our reconstructed SMB data show a slightly negative average value during the 1970s, even less negative in the 1980s, and then increasingly negative values in recent decades with an abrupt change in 2003 (Fig. 2 and 3). For this period (1967-2015), the year 2003 with its remarkable heatwave remains the most negative glacier-wide SMB year (-2.26 m.w.e. a^{-1} on average), with 1984 being the most positive year of the study period (+0.85 m.w.e. a^{-1} on average). The area-weighted average SMB is slightly less negative than the mean annual glacier-wide SMB, showing a light asymmetry in the probability distribution function (PDF) (Fig. 2c).

3.3   Regional and topographical trends

Here we analyse the main trends for the glacierized massifs and for some relevant topographical parameters. The reported glacier-wide SMBs are only area-weighted if specifically mentioned. Interesting differences appear once the dataset is divided
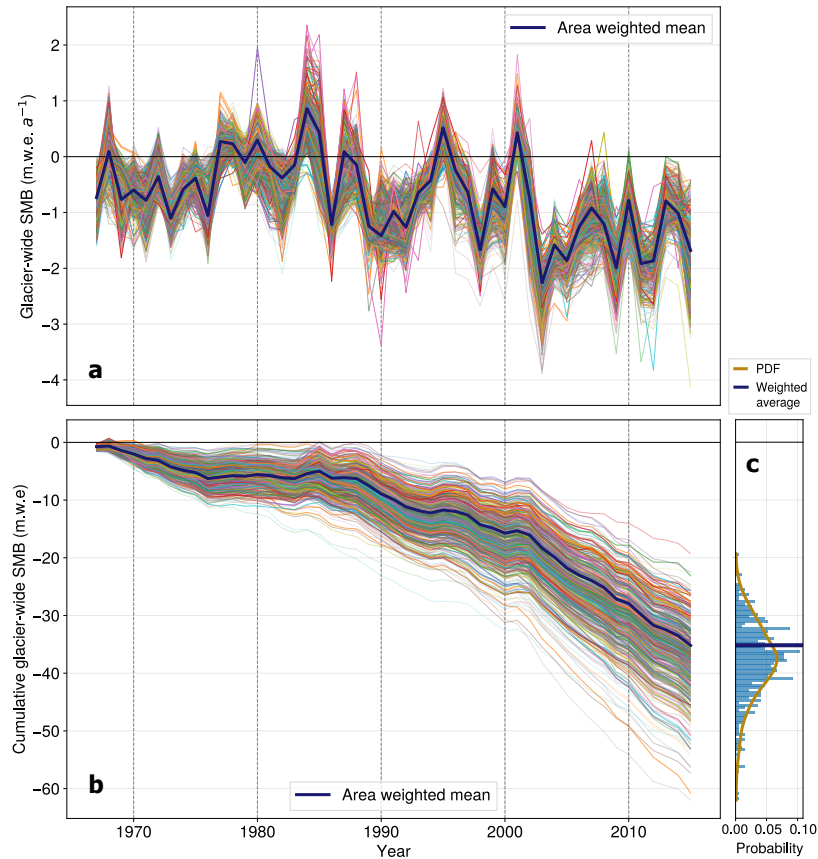
Figure 2. (a) Annual glacier-wide SMB reconstruction and (b) cumulative glacier-wide SMB reconstructions of all the glaciers in the French Alps ($N = 661$) between 1967 and 2015. For each individual glacier, line thickness depends on glacier area, with smaller glaciers having thinner lines. The histogram (c) indicates the distribution and probability density function (PDF) of the 1967-2015 cumulative SMB (m w.e.) of the dataset.

into mountain ranges (Fig. 4). The Mont-Blanc massif presents the lowest mass loss over the entire study period, with an average cumulative loss over the 1967-2015 period of 36.42 34.10 m.w.e. This is probably due to its northern location within the French Alps and its large high altitude accumulation areas, which resulted in more positive or less negative SMBs, especially during the 1980-2000s. Oisans is the massif with the second lowest average cumulative mass loss (38.35 37.20 m.w.e.). Its

5 glaciers have average altitudes ranging from 2290 to 3470 m.a.s.l., with around 50% of them having mean altitudes over 3000 m.a.s.l. and with about 40% of glaciers (including most of the large ones) having a northern aspect. Glaciers in Haute-Tarentaise present similar characteristics to those from Oisans, with mean altitudes ranging between 2300 and 3600 m.a.s.l., with about 60% of the glaciers above 3000 m.a.s.l. This less negative trend was especially important during the recent years with high mass losses from 2003 onwards. On the other hand, the Ubaye, Champsaur, Chablais and Haute-Maurienne massifs
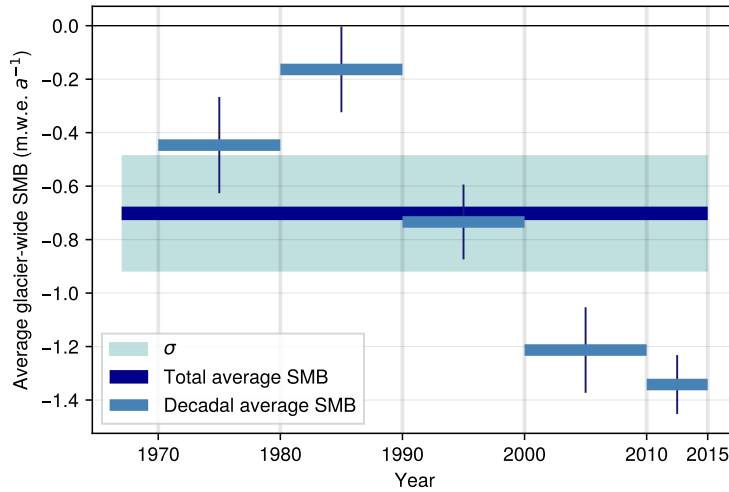
Figure 3. Averaged area-weighted decadal glacier-wide SMB for the French Alps with decadal uncertainties. The total area-weighted glacier-wide SMB is estimated for the 1967-2015 period

appear as the most affected mountain ranges with cumulative mass losses reaching between ~~44 and 45~~ 41 and 46 m.w.e. for the four massifs over the 1967-2015 period. The Chablais range has a very small number of glaciers remaining, all of them at rather low altitudes (2200-2900 m.a.s.l.), relatively small (0.01 - 1.1 km^2), and with a northwestern aspect. Despite being the northernmost mountain range in the French Alps, its low altitude is most likely the main reason for the very negative SMBs, which were under the regional average even during the positive years in the 1980s. The Champsaur range shows a similar situation, with very small glaciers (0.03 - 0.89 km^2) lying at relatively low altitudes (2300-3100 m.a.s.l.) in the southernmost latitudes of the Alps (44°7'). Finally, the situation of the Ubaye massif is quite similar to the one of Champsaur, being the southernmost glacierized massif in the French Alps, with a strong mediterranean influence. Such glaciers are remnants of the Little Ice Age, far from being in equilibrium with the warming climate, and can quickly lose a lot of mass through non-dynamic downwasting (Paul et al., 2004).

When classifying the SMB time series by glacier surface area, we encounter the following patterns, with n being the number of glaciers in the subset and s its standard deviation: (1) Very small glaciers (< 0.5 km^2; n = 534; $\overline{SMB}_{1967-2015}$ = ~~-0.82~~ -0.79 m.w.e. a^{-1}; s = ~~0.21~~ 0.23 m.w.e. a^{-1}) present more negative glacier-wide SMBs than (2) small/medium glaciers (ranging from 0.5 to 2 km^2; n = 93; $\overline{SMB}_{1967-2015}$ = -0.76 m.w.e. a^{-1}; s = ~~0.16~~ 0.17 m.w.e. a^{-1}) and (3) large glaciers (> 2 km^2; n = 34; $\overline{SMB}_{1967-2015}$ = ~~-0.72~~ -0.71 m.w.e. a^{-1}; s = 0.10 m.w.e. a^{-1})~~. (Fig. S8).~~ Very small glaciers present a larger spread of values than small/medium and large glaciers (s = ~~0.21~~ 0.23 m.w.e. a^{-1} versus 0.17 and 0.10 m.w.e. a^{-1}, respectively). As explained in Sect. 2, the uncertainties for very small glaciers are greater due to their ~~underrepresentation~~ under-representation in the training dataset, meaning that ~~analysis~~ analyses based on small glaciers have to be taken with greater care. The effects of these trends can be seen in the PDF of the cumulative SMB reconstructions

(Fig. 2c), where the area-weighted mean lies slightly outside the PDF maximum, showing how a great number of small glaciers are presenting higher losses. On the other hand, a clearer relationship between the glacier slope (computed here as the lowermost 20% altitudinal range slope) and glacier-wide SMB arises, with steeper glaciers having less negative glacier-wide SMBs (Fig. ~~S4 and S7~~ S6 and S9). Glaciers with a gentle tongue slope generally present longer response times and higher ice thickness, which are associated with more negative mass balances ~~(Hoelzle et al., 2003; Huss and Fischer, 2016)~~ (Hoelzle et al., 2003; Huss and Fischer, 2016; Zekollari et al., 2020). These results are in agreement with the findings by Fischer et al. (2015), who computed the geodetic mass balance of all the Swiss glaciers for the 1980-2010 period. Overall, the topographical relationships found here are similar, although more negative than for the Swiss Alps ~~(Huss, 2012; Huss and Hock, 2015)~~ (Huss, 2012; Huss et al., 2015), showing how the southernmost glaciers in the Écrins and Vanoise regions present stronger glacier mass losses. This is mostly due to their mediterranean climatic influence compared to the more continental Swiss and Austrian glaciers, which results in more negative SMB in a warming climate (Oerlemans and Reichert, 2000). Nonetheless, results from this type of bivariate analysis can show rather biased trends, since the topographical variables are highly intercorrelated, with for example small glaciers having steeper slopes and *vice versa* (Gardent et al., 2014). The position and evolution of the equilibrium line can totally reverse the trends of small or steep glaciers, so these relationships can strongly vary depending on the region or time period observed.

3.4 Comparison with previous studies and observations

In order to put into perspective the reconstructions presented in this study, we compare them to an updated version from the Marzeion et al. (2015) reconstructions (B. Marzeion, personal communication, October 2019 - January 2020), and to all the available glacier-wide SMB observations in the French Alps. The goal of this comparison is not to draw conclusions on the quality of either reconstruction, but to analyse the differences among them and to try to understand the causes. In the updated version of Marzeion et al. (2015) - referred as M_{15U} from now on - a global SMB model relying on temperature and solid precipitation was used to reconstruct SMB time series for all the glaciers in the world present in the Randolph Glacier Inventory (Consortium, 2017). This model was optimized based on ~~three~~ five parameters: the temperature sensitivity of the glacier ~~,~~ (local); and a precipitation correction factor~~and a bias correction~~, precipitation lapse rate, temperature threshold for solid precipitation and melt temperature threshold (global). As in Bolibar et al. (2020b), the approach by M_{15U} was cross-validated respecting the spatiotemporal independence in order to evaluate its performance for unobserved glaciers and years. ~~Prior to contrasting the results, three important different aspects between our approach and the one of M_{15U} 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 use 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} force their updated model with CRU 6.0 (update of Harris et al., 2014), with 0.5° 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~~
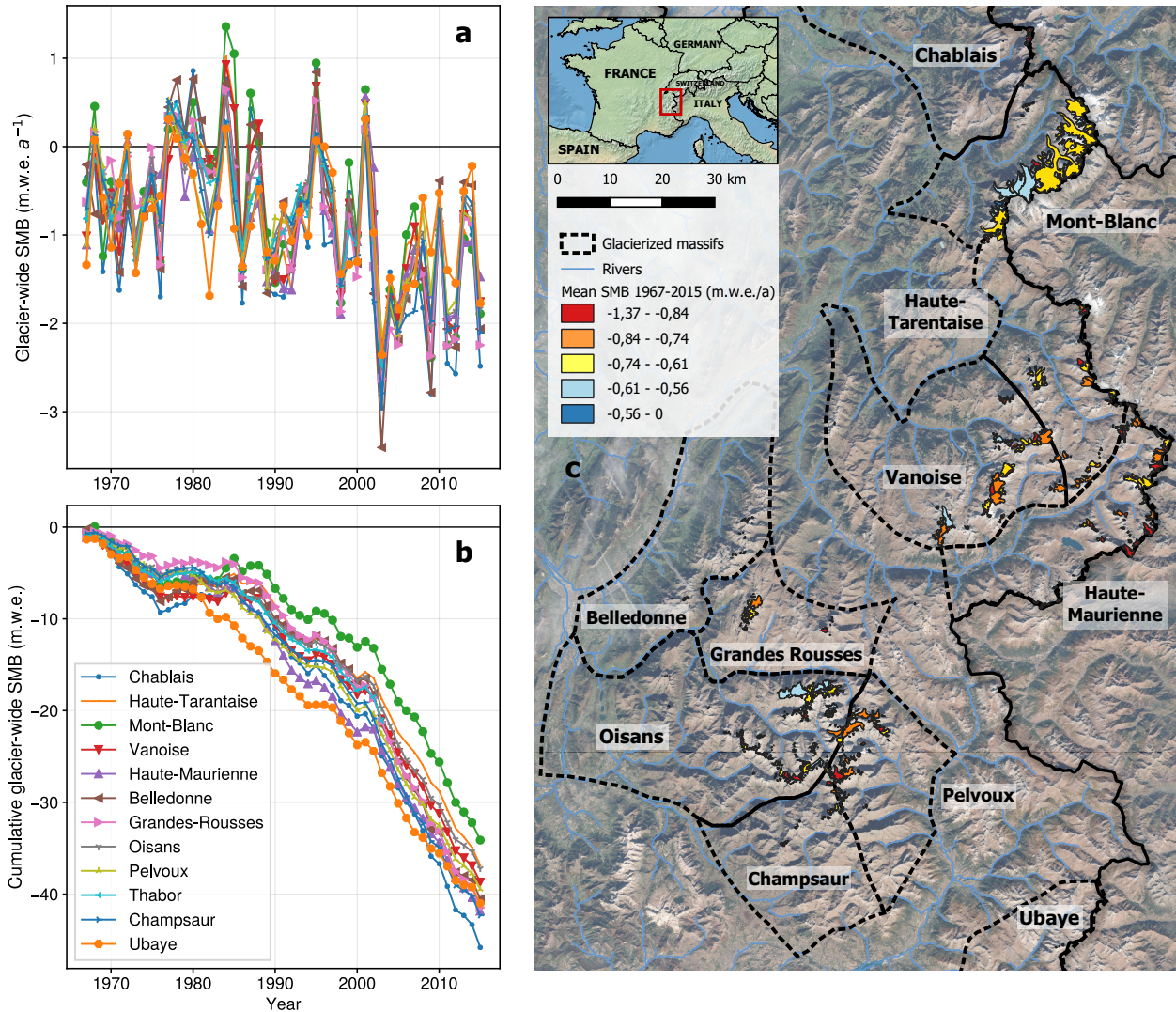
Figure 4. (a) Averaged annual glacier-wide SMB and (b) cumulative averaged glacier-wide SMB time series for each of the massifs in the French Alps between 1967 and 2015. (c) Glacierized massifs in the French Alps with the average glacier-wide SMB for the 1967-2015 period. Coordinates of bottom left map corner: 44º32' N, 5º40' E. Coordinates of the top right map corner: 46º08' N, 7º17' E. (Basemap © imagico.de)

high-altitude stations are assimilated. The cross-validations of both studies determined a performance with an average RMSE of 0.66 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 m.w.e. a^{-1} and an r^2 of 0.79 for this study. However, due to the 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. All the specific differences and details between the two models can be found in Sect. S2 from the supplement.

As shown in Figure 5, the interannual variability, driven by climate, is quite similar between the two reconstructions. Conversely, important differences are found for different subperiods in the amplitude of the area-weighted mean glacier-wide SMB series. These differences are the greatest in the 1970s, 1980s and 2010s, with similar average values for the 1990s and 2000s (Fig. 5 and ~~S5~~S7). M_{15U} presents less negative and more positive glacier-wide SMB values in the 1970s, but on the contrary, it presents more negative values in the 1980s compared to our results. We believe there might be two potential reasons for this: (1) In 1976 there was a shift in the winter mass balance regime in the French Alps, with more humid winters bringing more accumulation; and in 1982 there was a shift in the summer mass balance, resulting in increased ablation (Thibert et al., 2013). Since both models use parameterized or statistical relationships for SMB response to precipitation and temperature, they are likely to react differently to these changes. A similar situation is found from the year 2003 onwards, where there was a substantial increase in temperatures and mass loss (e.g. Six and Vincent, 2014). Our reconstructions show a marked change in 2003 (change of slope in the cumulative plot in Fig. 5), whereas M_{15U} present a rather linear trend. 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. (2) For the 1967-1983 interval, the amount of available glacier-wide SMB data for training is much lower than for the rest of the period (green numbers in Fig. 5). This is likely the reason why the differences between our reconstructions and observations are greater for that period (Fig. 5). On the other hand, the similarities between our reconstructions and the observations for the 1984-2014 period are explained by the fact that the 32 glaciers with observations represent around 45% of the total glacierized area in the French Alps in the year 2003. For the periods before and after this interval, differences and uncertainties in the reconstructed values are greater because of the smaller sample size.

Important similarities between observations and the reconstructed glacier-wide SMB values for the 1984-2015 period in this study (Fig. 5) question a possible overfitting of the reconstructions to the training data. First, for the vast majority of the 661 French glaciers, the reconstructions are based on an ensemble of cross-validated models, which intrinsically limits overfitting (see Sect. 2). Second, we analysed the deviation to the climatological mass-balance signal of the SMB for each cluster of glacier-sizes. This analysis is presented in the ~~Supplementary~~supplementary material. It reveals that the similarities between observations and the reconstructed glacier-wide SMB values for the 1984-2015 period in Fig. 5 proceed from big glaciers, that dominate both in the area-weighted reconstructions and in the observations ~~.~~(Fig. S3 and S4). However, for the other glacier-size classes, our reconstruction shows different patterns from the data in the observations, which suggests that the model is not overfitting (Fig. ~~S1~~S3).

4 Data availability

The full glacier-wide SMB dataset and the detailed topographical information of all the French alpine glaciers is available in the following Zenodo repository: https://doi.org/10.5281/zenodo.3663630 (Bolibar et al., 2020a).
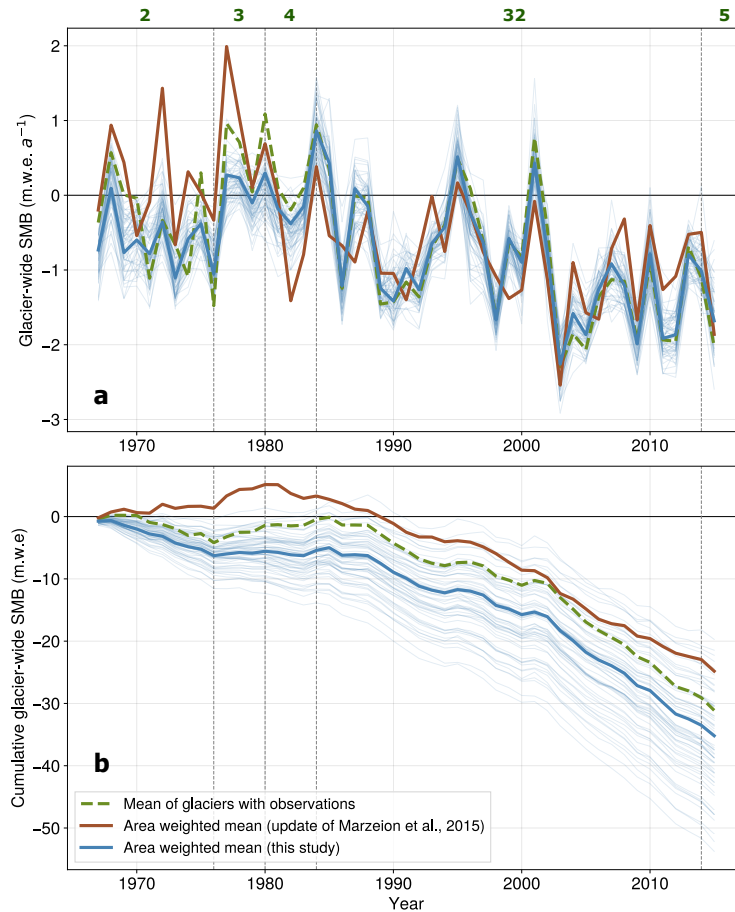
Figure 5. Comparison of (a) annual and (b) cumulative glacier-wide SMB simulations in the French Alps between this study, reconstructions from an update from Marzeion et al. (2015) and the mean of all observations available in the French Alps. Green numbers indicate the number of glaciers with SMB observations for each period and thin light blue lines indicate the area-weighted mean of each of the cross-validation ensemble members.

5 Conclusions

We presented a dataset of annual glacier-wide SMB of all the glaciers in the French Alps (44° - 46°13'N, 5.08° - 7.67°E) for the 1967-2015 period (Bolibar et al., 2020a). This dataset has been reconstructed using deep learning (i.e. an artificial neural network), based on direct and remote sensing annual glacier-wide SMB observations, climate reanalysis and topographical data from multitemporal glacier inventories. The deep learning model is capable of reconstructing glacier-wide SMB time series for unobserved glaciers in the same region based on patterns and structures learnt by the artificial neural network from the observations and their relationships with predictors. An extensive cross-validation was implemented to understand the characteristics of the SMB signal in the region and to assess the method's validity and uncertainty. The average accuracy

(RMSE) of the dataset is estimated at ~~0.49~~ 0.55 m.w.e. a^{-1} with an explained variance (r^2) of ~~79~~75%. Reconstructions show a mean area-weighted glacier-wide SMB of ~~-0.72~~ -0.71±~~0.20~~ 0.21 (σ) m.w.e. a^{-1} for the 1967-2015 period. Important differences are found among different massifs, with the Mont-Blanc (~~-0.74~~ -0.69 m.w.e. a^{-1}), Oisans and Haute-Tarentaise ranges (~~-0.78~~ -0.75 m.w.e. a^{-1} both) presenting the lowest mass losses and the Chablais (~~-0.90~~ -0.93 m.w.e. a^{-1}), ~~Ubaye and Champsaur~~ ~~(-0.91~~ Champsaur and Haute-Maurienne (-0.86 m.w.e. a^{-1} both) and Ubaye (-0.83 m.w.e. a^{-1}) showing the highest losses. In order to put these results into perspective, this reconstruction was compared to all available glacier-wide SMB observations in the French Alps as well as the physical/empirical reconstructions from another study (update from Marzeion et al., 2015). Interesting differences were found between the two methods, highlighting the different sensitivities and responses of different approaches to climate shifts that occurred during the study period. These differences are particularly relevant in the 1970s and 1980s, previous to a winter precipitation and summer temperature shift that occurred in the French Alps in the years 1976 and 1982, respectively. Moreover, after the famous 2003 European heatwave, glaciers experienced an acceleration in mass loss which is well captured by our reconstruction. This open glacier-wide SMB dataset can be useful for hydrological or ecological studies in need of ~~meltwater~~ net glacier mass contributions of glacierized catchments in the French Alps. Moreover, the publication of such open datasets is the cornerstone of future community-based data-driven scientific studies.

6   Code availability

The source code of ALPGM v1.1 is accessible at https://github.com/JordiBolibar/ALPGM, with its DOI: 10.5281/zenodo.3609136 (Bolibar, 2020).

References

Benn, D. I. and Evans, D. J. A.: Glaciers & glaciation, Routledge, New York, NY, USA, 2nd edn., http://www.imperial.eblib.com/EBLWeb/patron/?target=patron&extendedid=P_615876_0, oCLC: 878863282, 2014.

Berthier, E., Vincent, C., Magnússon, E., Gunnlaugsson, , Pitte, P., Le Meur, E., Masiokas, M., Ruiz, L., Pálsson, F., Belart, J. M. C., and Wagnon, P.: Glacier topography and elevation changes derived from Pléiades sub-meter stereo images, The Cryosphere, 8, 2275–2291, https://doi.org/10.5194/tc-8-2275-2014, https://www.the-cryosphere.net/8/2275/2014/, 2014.

Berthier, E., Cabot, V., Vincent, C., and Six, D.: Decadal Region-Wide and Glacier-Wide Mass Balances Derived from Multi-Temporal ASTER Satellite Digital Elevation Models. Validation over the Mont-Blanc Area, Frontiers in Earth Science, 4, https://doi.org/10.3389/feart.2016.00063, http://journal.frontiersin.org/Article/10.3389/feart.2016.00063/abstract, 2016.

Bolibar, J.: ALPGM (ALpine Parameterized Glacier Model) v1.1, https://doi.org/10.5281/zenodo.3609136, https://doi.org/10.5281/zenodo.3609136, 2020.

Bolibar, J., Rabatel, A., Gouttevin, I., and Galiez, C.: A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967-2015, Zenodo, https://doi.org/10.5281/zenodo.3663630, https://doi.org/10.5281/zenodo.3663630, [Dataset], 2020a.

Bolibar, J., Rabatel, A., Gouttevin, I., Galiez, C., Condom, T., and Sauquet, E.: Deep learning applied to glacier evolution modelling, The Cryosphere, 14, 565–584, https://doi.org/10.5194/tc-14-565-2020, https://www.the-cryosphere.net/14/565/2020/, 2020b.

Brun, F., Berthier, E., Wagnon, P., Kääb, A., and Treichler, D.: A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016, Nature Geoscience, 10, 668–673, https://doi.org/10.1038/ngeo2999, http://www.nature.com/doifinder/10.1038/ngeo2999, 2017.

Cauvy-Frauníe, S. and Dangles, O.: A global synthesis of biodiversity responses to glacier retreat, Nature Ecology & Evolution, 3, 1675–1685, https://doi.org/10.1038/s41559-019-1042-8, http://www.nature.com/articles/s41559-019-1042-8, 2019.

Consortium, R. G. I.: Randolph Glacier Inventory 6.0, https://doi.org/10.7265/N5-RGI-60, http://www.glims.org/RGI/randolph60.html, type: dataset, 2017.

Davaze, L., Rabatel, A., Dufour, A., Arnaud, Y., and Hugonnet, R.: Region-wide annual glacier surface mass balance for the European Alps from 2000 to 2016, Frontiers in Earth Science, 8, 149, publisher: Frontiers, 2020.

Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B., and Mérindol, L.: Reanalysis of 44 Yr of Climate in the French Alps (1958–2002): Methodology, Model Validation, Climatology, and Trends for Air Temperature and Precipitation, Journal of Applied Meteorology and Climatology, 48, 429–449, https://doi.org/10.1175/2008JAMC1808.1, http://journals.ametsoc.org/doi/abs/10.1175/2008JAMC1808.1, 2009.

Dussaillant, I., Berthier, E., Brun, F., Masiokas, M., Hugonnet, R., Favier, V., Rabatel, A., Pitte, P., and Ruiz, L.: Two decades of glacier mass loss along the Andes, Nature Geoscience, https://doi.org/10.1038/s41561-019-0432-5, http://www.nature.com/articles/s41561-019-0432-5, 2019.

Fausett, L. V.: Fundamentals of neural networks: architectures, algorithms, and applications, Prentice Hall, Englewood Cliffs, N.J., oCLC: 28215780, 1994.

Fischer, M., Huss, M., and Hoelzle, M.: Surface elevation and mass changes of all Swiss glaciers 1980–2010, The Cryosphere, 9, 525–540, https://doi.org/10.5194/tc-9-525-2015, https://www.the-cryosphere.net/9/525/2015/, 2015.

Gardent, M., Rabatel, A., Dedieu, J.-P., and Deline, P.: Multitemporal glacier inventory of the French Alps from the late 1960s to the late 2000s, Global and Planetary Change, 120, 24–37, https://doi.org/10.1016/j.gloplacha.2014.05.004, https://linkinghub.elsevier.com/retrieve/pii/S092181811400099X, 2014.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset: UPDATED HIGH-RESOLUTION GRIDS OF MONTHLY CLIMATIC OBSERVATIONS, International Journal of Climatology, 34, 623–642, https://doi.org/10.1002/joc.3711, http://doi.wiley.com/10.1002/joc.3711, 2014.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning, Springer Series in Statistics, Springer New York, New York, NY, https://doi.org/10.1007/978-0-387-84858-7, http://link.springer.com/10.1007/978-0-387-84858-7, 2009.

Hock, R., Bliss, A., Marzeion, B., Giesen, R. H., Hirabayashi, Y., Huss, M., Radić, V., and Slangen, A. B. A.: GlacierMIP – A model intercomparison of global-scale glacier mass-balance models and projections, Journal of Glaciology, 65, 453–467, https://doi.org/10.1017/jog.2019.22, https://www.cambridge.org/core/product/identifier/S0022143019000224/type/journal_article, 2019.

Hoelzle, M., Haeberli, W., Dischl, M., and Peschke, W.: Secular glacier mass balances derived from cumulative glacier length changes, Global and Planetary Change, 36, 295–306, https://doi.org/10.1016/S0921-8181(02)00223-0, https://linkinghub.elsevier.com/retrieve/pii/S0921818102002230, 2003.

Huss, M.: Extrapolating glacier mass balance to the mountain-range scale: the European Alps 1900–2100, The Cryosphere, 6, 713–727, https://doi.org/10.5194/tc-6-713-2012, https://www.the-cryosphere.net/6/713/2012/, 2012.

Huss, M. and Fischer, M.: Sensitivity of Very Small Glaciers in the Swiss Alps to Future Climate Change, Frontiers in Earth Science, 4, https://doi.org/10.3389/feart.2016.00034, http://journal.frontiersin.org/Article/10.3389/feart.2016.00034/abstract, 2016.

Huss, M. and Hock, R.: A new model for global glacier change and sea-level rise, Frontiers in Earth Science, 3, https://doi.org/10.3389/feart.2015.00054, http://journal.frontiersin.org/Article/10.3389/feart.2015.00054/abstract, 2015.

Huss, M. and Hock, R.: Global-scale hydrological response to future glacier mass loss, Nature Climate Change, 8, 135–140, https://doi.org/10.1038/s41558-017-0049-x, http://www.nature.com/articles/s41558-017-0049-x, 2018.

Huss, M., Dhulst, L., and Bauder, A.: New long-term mass-balance series for the Swiss Alps, Journal of Glaciology, 61, 551–562, https://doi.org/10.3189/2015JoG15J015, https://www.cambridge.org/core/product/identifier/S0022143000204036/type/journal_article, 2015.

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B. J., Elmore, A. C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J. S., Koppes, M., Kraaijenbrink, P. D. A., Kulkarni, A. V., Mayewski, P. A., Nepal, S., Pacheco, P., Painter, T. H., Pellicciotti, F., Rajaram, H., Rupper, S., Sinisalo, A., Shrestha, A. B., Viviroli, D., Wada, Y., Xiao, C., Yao, T., and Baillie, J. E. M.: Importance and vulnerability of the world's water towers, Nature, 577, 364–369, https://doi.org/10.1038/s41586-019-1822-y, http://www.nature.com/articles/s41586-019-1822-y, 2020.

Ingrassia, S. and Morlini, I.: Neural Network Modeling for Small Datasets, Technometrics, 47, 297–311, https://doi.org/10.1198/004017005000000058, http://www.tandfonline.com/doi/abs/10.1198/004017005000000058, 2005.

Krogh, A. and Vedelsby, J.: Neural network ensembles, cross validation, and active learning, in: Advances in neural information processing systems, pp. 231–238, 1995.

Kääb, A., Berthier, E., Nuth, C., Gardelle, J., and Arnaud, Y.: Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas, Nature, 488, 495–498, https://doi.org/10.1038/nature11324, http://www.nature.com/articles/nature11324, 2012.

Marzeion, B., Jarosch, A. H., and Hofer, M.: Past and future sea-level change from the surface mass balance of glaciers, The Cryosphere, 6, 1295–1322, https://doi.org/10.5194/tc-6-1295-2012, https://www.the-cryosphere.net/6/1295/2012/, 2012.

Marzeion, B., Leclercq, P. W., Cogley, J. G., and Jarosch, A. H.: Brief Communication: Global reconstructions of glacier mass change during the 20th century are consistent, The Cryosphere, 9, 2399–2404, https://doi.org/10.5194/tc-9-2399-2015, https://www.the-cryosphere.net/9/2399/2015/, 2015.

Oerlemans, J. and Reichert, B. K.: Relating glacier mass balance to meteorological data by using a seasonal sensitivity characteristic, Journal of Glaciology, 46, 1–6, https://doi.org/10.3189/172756500781833269, https://www.cambridge.org/core/product/identifier/S0022143000213786/type/journal_article, 2000.

Paul, F., Kääb, A., Maisch, M., Kellenberger, T., and Haeberli, W.: Rapid disintegration of Alpine glaciers observed with satellite data: DISINTEGRATION OF ALPINE GLACIERS, Geophysical Research Letters, 31, n/a–n/a, https://doi.org/10.1029/2004GL020816, http://doi.wiley.com/10.1029/2004GL020816, 2004.

Rabatel, A., Dedieu, J.-P., and Vincent, C.: Using remote-sensing data to determine equilibrium-line altitude and mass-balance time series: validation on three French glaciers, 1994–2002, Journal of Glaciology, 51, 539–546, https://doi.org/10.3189/172756505781829106, https://www.cambridge.org/core/product/identifier/S0022143000210733/type/journal_article, 2005.

Rabatel, A., Dedieu, J. P., and Vincent, C.: Spatio-temporal changes in glacier-wide mass balance quantified by optical remote sensing on 30 glaciers in the French Alps for the period 1983–2014, Journal of Glaciology, 62, 1153–1166, https://doi.org/10.1017/jog.2016.113, https://www.cambridge.org/core/product/identifier/S0022143016001131/type/journal_article, 2016.

Rastner, P., Prinz, R., Notarnicola, C., Nicholson, L., Sailer, R., Schwaizer, G., and Paul, F.: On the Automated Mapping of Snow Cover on Glaciers and Calculation of Snow Line Altitudes from Multi-Temporal Landsat Data, Remote Sensing, 11, 1410, https://doi.org/10.3390/rs11121410, https://www.mdpi.com/2072-4292/11/12/1410, 2019.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography, 40, 913–929, https://doi.org/10.1111/ecog.02881, http://doi.wiley.com/10.1111/ecog.02881, 2017.

Six, D. and Vincent, C.: Sensitivity of mass balance and equilibrium-line altitude to climate change in the French Alps, Journal of Glaciology, 60, 867–878, https://doi.org/10.3189/2014JoG14J014, https://www.cambridge.org/core/product/identifier/S0022143000202128/type/journal_article, 2014.

Thibert, E., Eckert, N., and Vincent, C.: Climatic drivers of seasonal glacier mass balances: an analysis of 6 decades at Glacier de Sarennes (French Alps), The Cryosphere, 7, 47–66, https://doi.org/10.5194/tc-7-47-2013, http://www.the-cryosphere.net/7/47/2013/, 2013.

Vincent, C., Fischer, A., Mayer, C., Bauder, A., Galos, S. P., Funk, M., Thibert, E., Six, D., Braun, L., and Huss, M.: Common climatic signal from glaciers in the European Alps over the last 50 years: Common Climatic Signal in the Alps, Geophysical Research Letters, 44, 1376–1383, https://doi.org/10.1002/2016GL072094, http://doi.wiley.com/10.1002/2016GL072094, 2017.

Vincent, C., Soruco, A., Azam, M. F., Basantes-Serrano, R., Jackson, M., Kjøllmoen, B., Thibert, E., Wagnon, P., Six, D., Rabatel, A., Ramanathan, A., Berthier, E., Cusicanqui, D., Vincent, P., and Mandal, A.: A Nonlinear Statistical Model for Extracting a Climatic Signal From Glacier Mass Balance Measurements, Journal of Geophysical Research: Earth Surface, 123, 2228–2242, https://doi.org/10.1029/2018JF004702, https://onlinelibrary.wiley.com/doi/abs/10.1029/2018JF004702, 2018.

Zekollari, H., Huss, M., and Farinotti, D.: Modelling the future evolution of glaciers in the European Alps under the EURO-CORDEX RCM ensemble, The Cryosphere, 13, 1125–1146, https://doi.org/10.5194/tc-13-1125-2019, https://www.the-cryosphere.net/13/1125/2019/, 2019.

Zekollari, H., Huss, M., and Farinotti, D.: On the Imbalance and Response Time of Glaciers in the European Alps, Geophysical Research Letters, 47, https://doi.org/10.1029/2019GL085578, https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085578, 2020.

16

Supplementary material: A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967-2015

Jordi Bolibar[1,2], Antoine Rabatel[1], Isabelle Gouttevin[3], and Clovis Galiez[4]

[1]Univ. Grenoble Alpes, CNRS, IRD, G-INP, Institut des Géosciences de l'Environnement (IGE, UMR 5001), Grenoble, France
[2]INRAE, UR RiverLy, Lyon-Villeurbanne, France
[3]Univ. Grenoble Alpes, Université de Toulouse, Météo-France, CNRS, CNRM, Centre d'Études de la Neige, Grenoble, France
[4]Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, Grenoble, France

Correspondence: Jordi Bolibar (jordi.bolibar@univ-grenoble-alpes.fr)

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.

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 M_{15U} need to be highlighted:
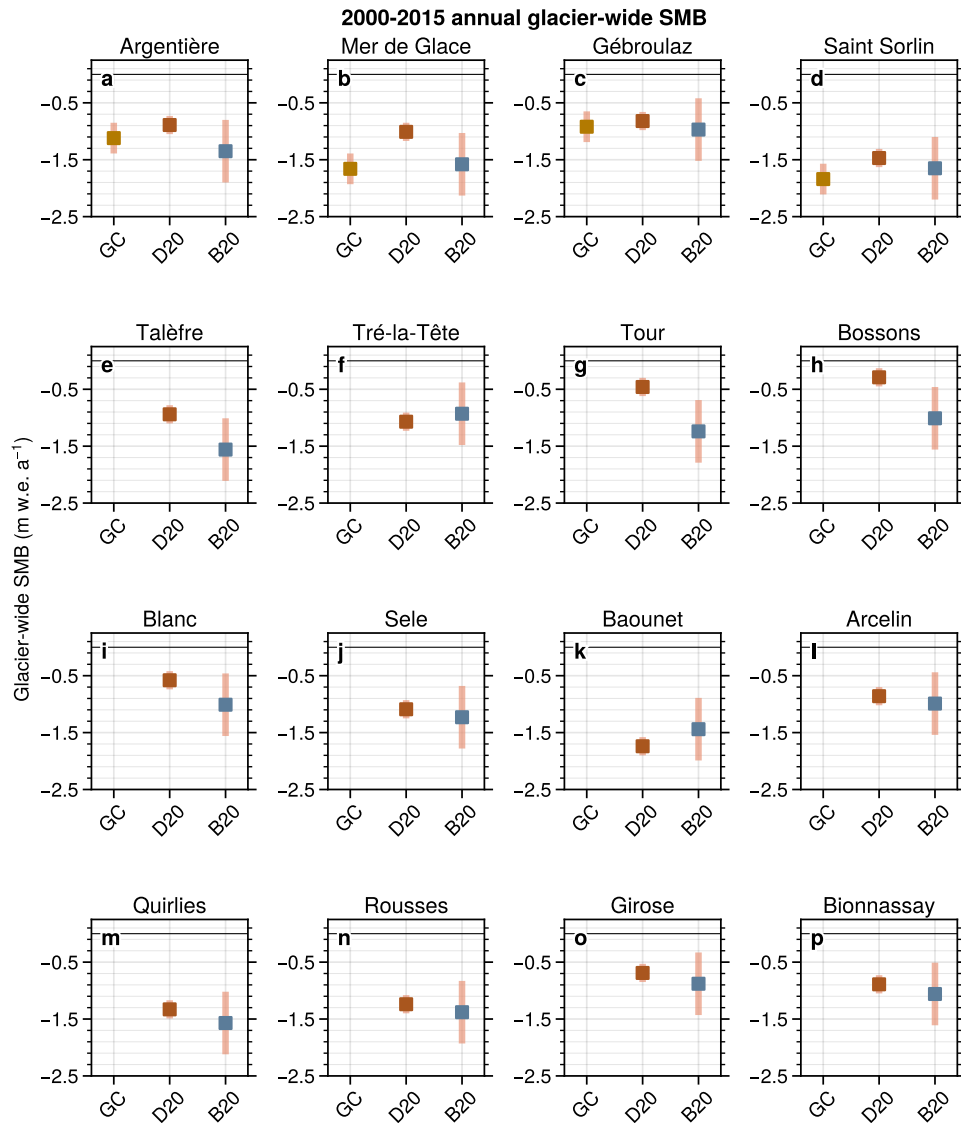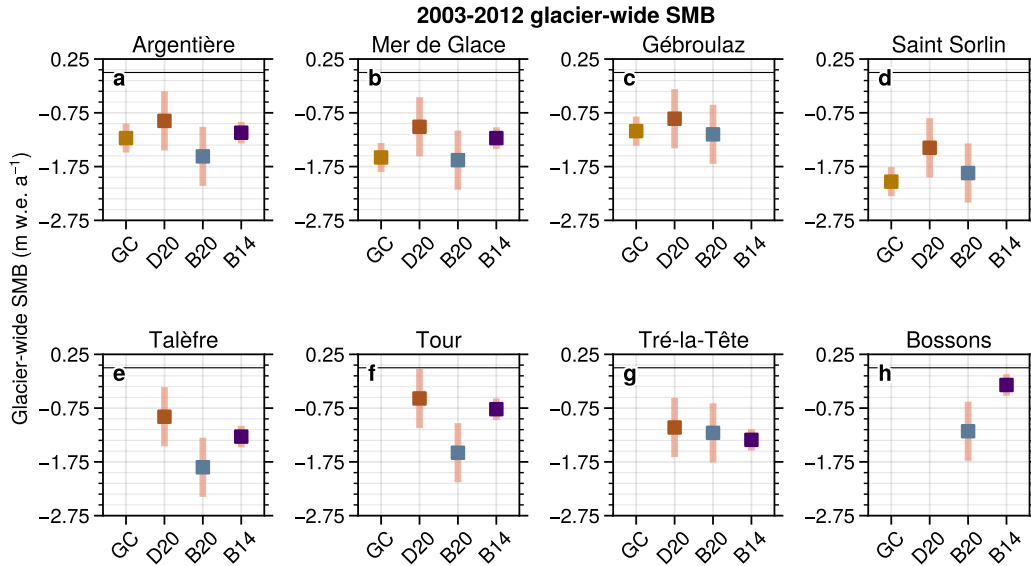
Figure S1. Comparison between glaciological observations from the GLACIOCLIM observatory (GC), ASTER geodetic mass balances from Davaze et al. (2020) (D20) and the deep learning reconstructions from the present study (B20).

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.

2003-2012 glacier-wide SMB

Figure S2. Comparison between glaciological observations from the GLACIOCLIM observatory (GC), ASTER geodetic mass balances from Davaze et al. (2020) (D20), the deep learning reconstructions from the present study (B20) and Pléiades geodetic mass balances from Berthier et al. (2014) (B14).

2. M_{15U} calibrated their model with global SMB observations, including 38 glaciers in the European Alps, 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° 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 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 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.

3 Influence of area in glacier-wide SMB signal and proof on non overfitting

Due to similarities between the averaged reconstructed glacier-wide surface mass balance (SMB) and the observations during the 1984-2015 period, we decided to include an analysis to isolate the topographical influence in the glacier-wide SMB signal,

in order to verify that the model is not overfitting. Since the climate signal is the main common driver of interannual variability of glacier-wide SMB in the region, one needs to find a way to isolate the topographical signal. In Fig. S3, the median reconstructed annual glacier-wide SMB of the 661 glaciers in the French Alps (i.e. the interannual variability, hence a proxy of the climate signal) is subtracted to the mean annual values of the observations and of 4 subsets of glaciers divided by area classes. Therefore, one can observe the residual influence of glacier area on the glacier-wide SMB signal. The influence of area on glaciers with observations is quite similar to glaciers with areas greater than 2 km^2, which is reasonable since glaciers with observations have an average of 4 km^2 (range: 0.3-31.8 km^2 in 2003). Moreover, one can see that even for a relatively short period of 30 years, the differences between the reconstructions for very small glaciers ($< 0.5\ km^2$) and observations are quite important, accounting for an average cumulative loss of more than 5 m.w.e. As stated in Sect. 2, this does not necessarily mean that the model has fully captured the topographical influence in the glacier-wide SMB signal in the region, but it does prove that the model is not overfitting since it exhibits consistent variations in SMB when the topographical predictors move away from the training data. Moreover, this is coherent with the importance attributed to topographical predictors (Bolibar et al., 2020).

The same analysis has been performed with the reconstructions from the updated version of Marzeion et al. (2015), shown in Fig. S4. The gradient with respect to glacier surface area appears to be similar, except for the behaviour of glaciers after 2007. Small and middle sized glaciers (0.1 - 2 km^2) switch to a positive influence, as opposite to large glaciers ($> 2\ km^2$), which transition to a negative influence. Conversely, our results show a more continuous trend, without a change of behaviour in the last years of the analysed period.
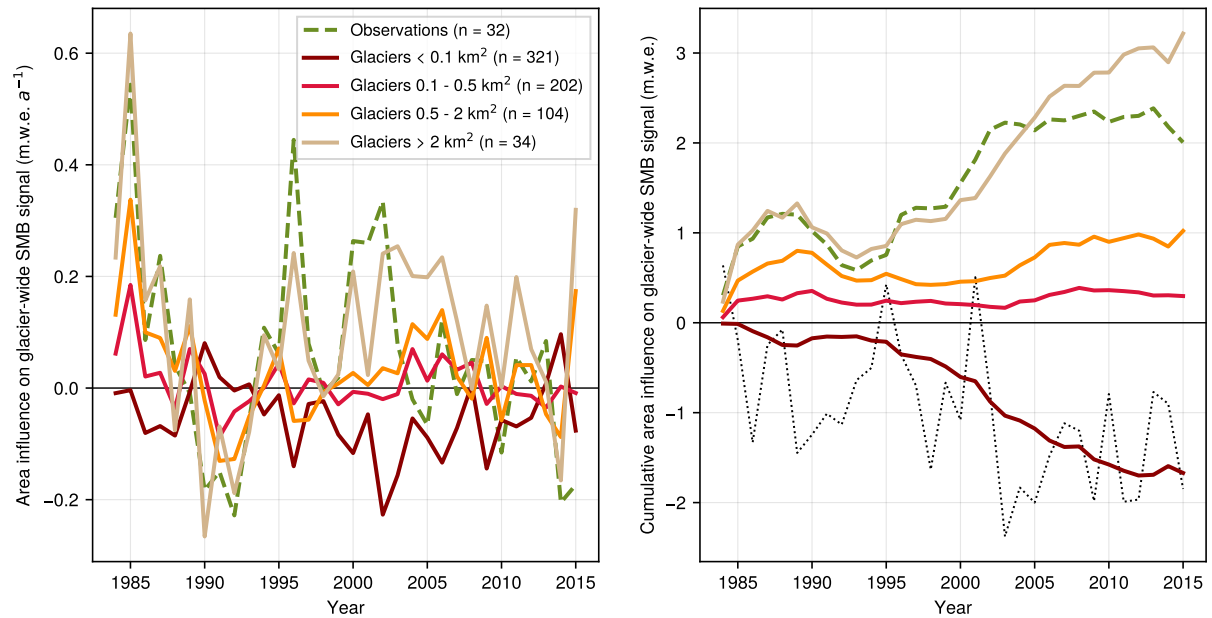
Figure S3. Influence of glacier area on the glacier-wide SMB signal. The reconstructed median annual glacier-wide SMB of the 661 glaciers in the French Alps can be seen as a proxy of the climate signal in the region. It is subtracted to the mean annual glacier-wide SMB of the glaciers with observations and to four different subsets of reconstructions divided into glacier area size, showing only the annual differences based on glacier area classes. The dotted line depicts the subtracted signal (non cumulative) in order to give some context.
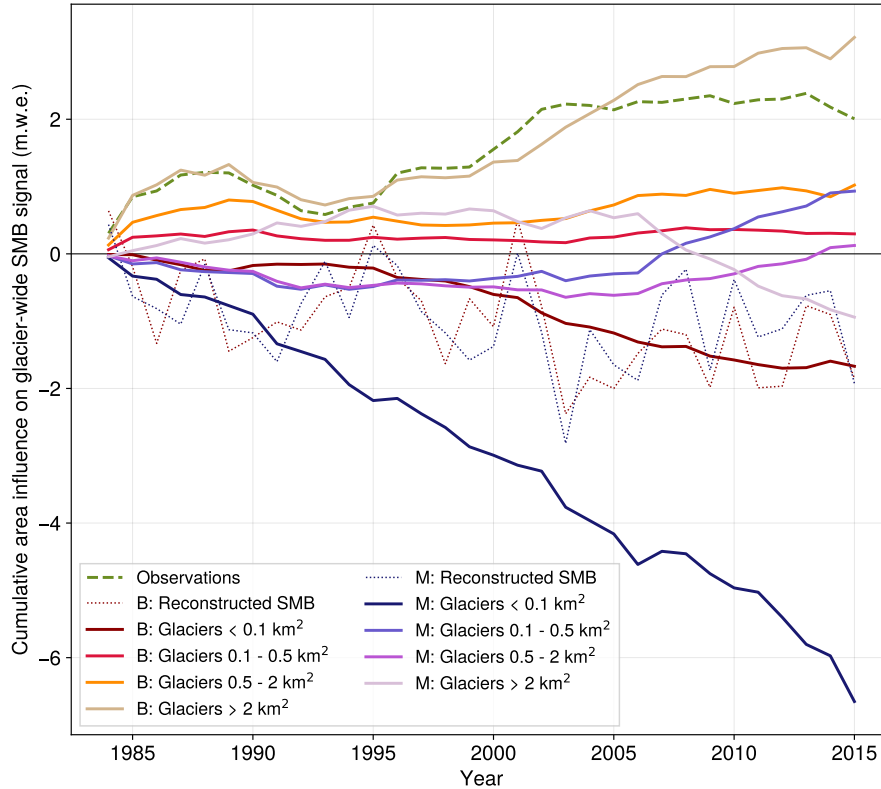
Figure S4. Same as S3 but comparing this study to the updated version of Marzeion et al. (2015). In the legend, "B" stands for Bolibar et al. (this study) and "M" for the update of Marzeion et al. (2015). Both models show a relatively similar gradient effect with respect to glacier area, with differences in the amplitude of the effects. The main differences appear from 2007, where small and middle sized glaciers (0.1 - 2 km^2) from the update of Marzeion et al. (2015) switch to a positive influence, as opposite to large glaciers (> 2 km^2), which transition to a negative influence. The reconstructed SMB dotted lines are not cumulative and they are depicted in order to give some context of the subtracted climate signal.
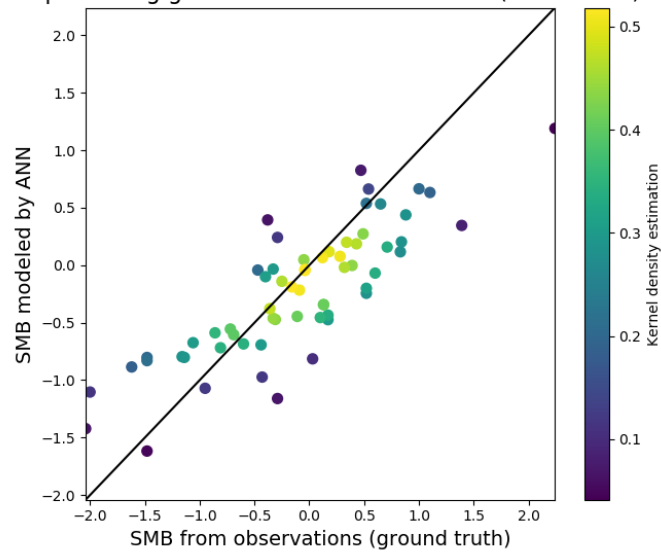
Figure S5. Cross-validation for annual glacier-wide SMB values outside the main 1984-2014 training period. The black line indicates the one-to-one reference. Simulations have been done from 1959, the earliest date with observations to validate against the maximum number of values. This serves to confirm that the model is capable of reproducing glacier-wide SMB outside the main observed period.
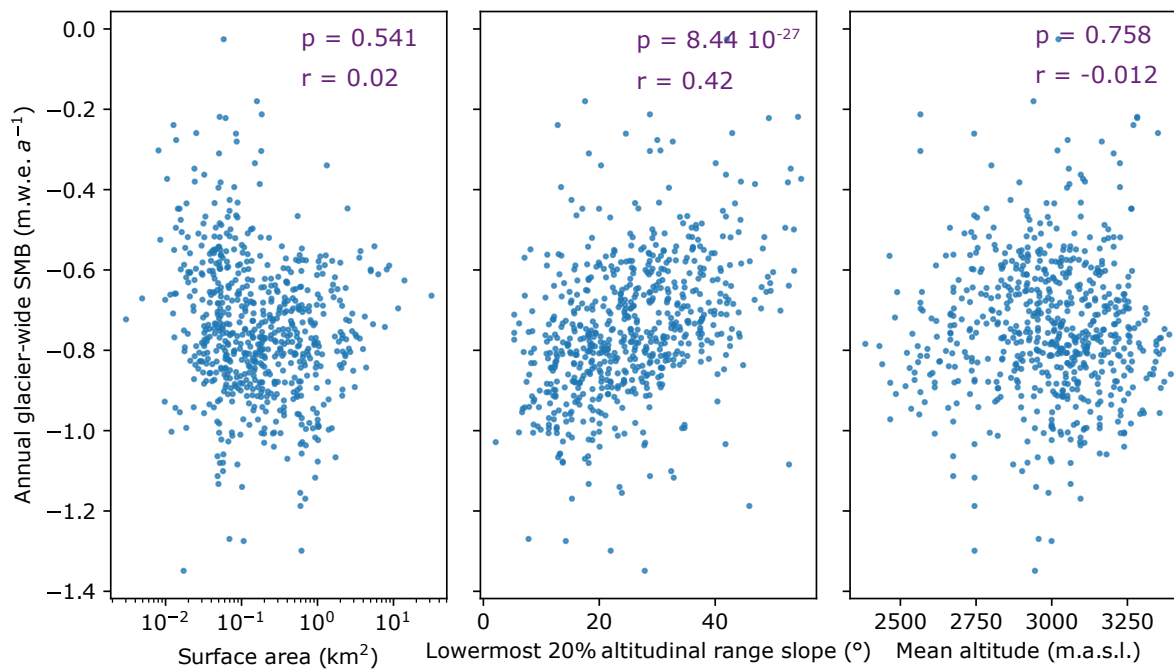
Figure S6. Average annual glacier-wide SMB for each glacier over the entire study period with respect to (a) glacier surface area, (b) the lowermost 20% altitudinal range slope and (c) mean glacier altitude. p indicates the p-value and r the correlation between the topographical variables and the average glacier-wide SMB.
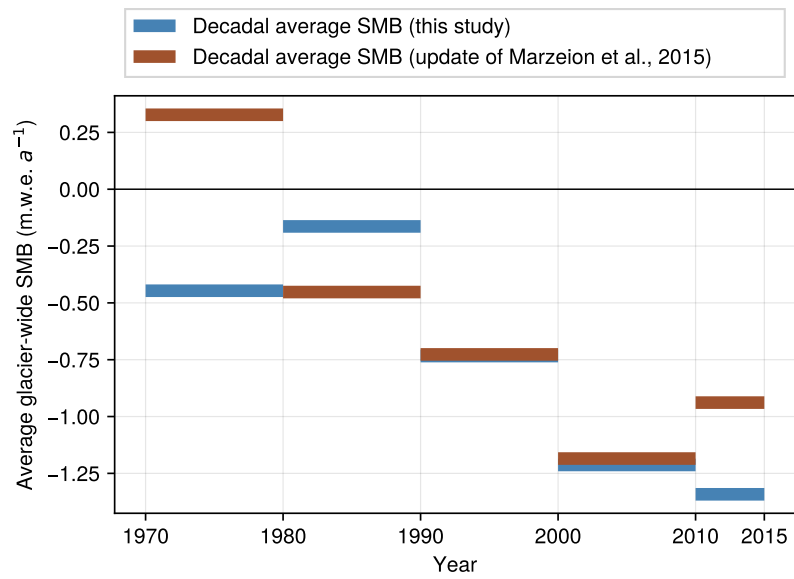
Figure S7. Comparison of area-weighted decadal glacier-wide SMB simulations in the French Alps between this study and an update from Marzeion et al. (2015).
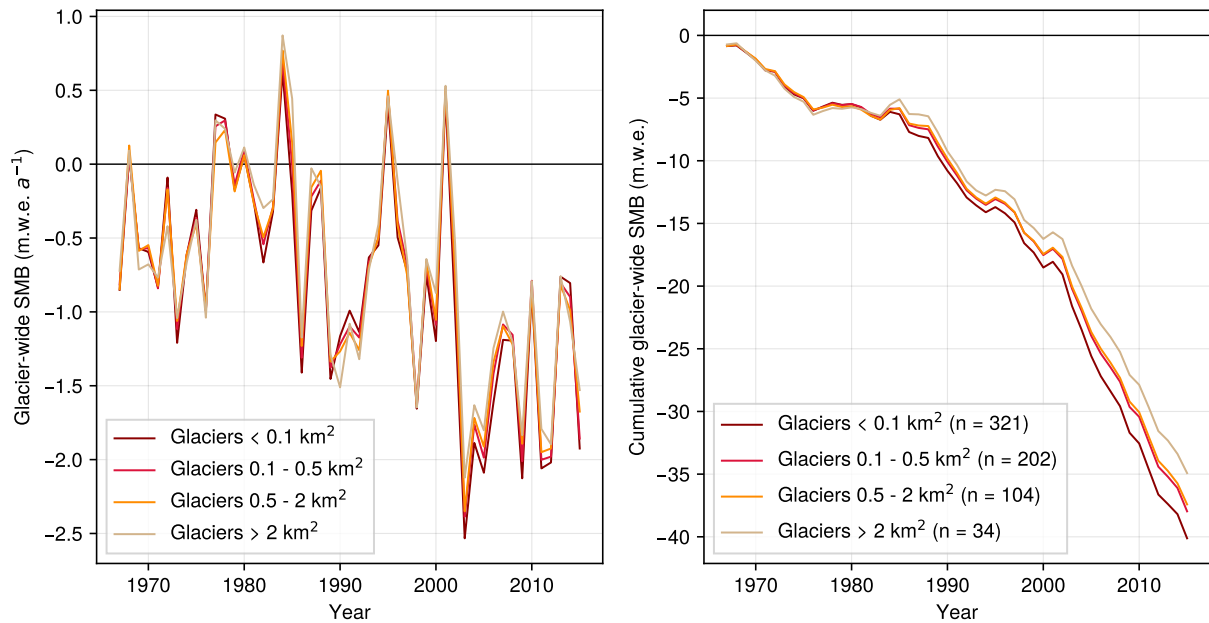

Figure S8. Average annual glacier-wide SMB per glacier area classes
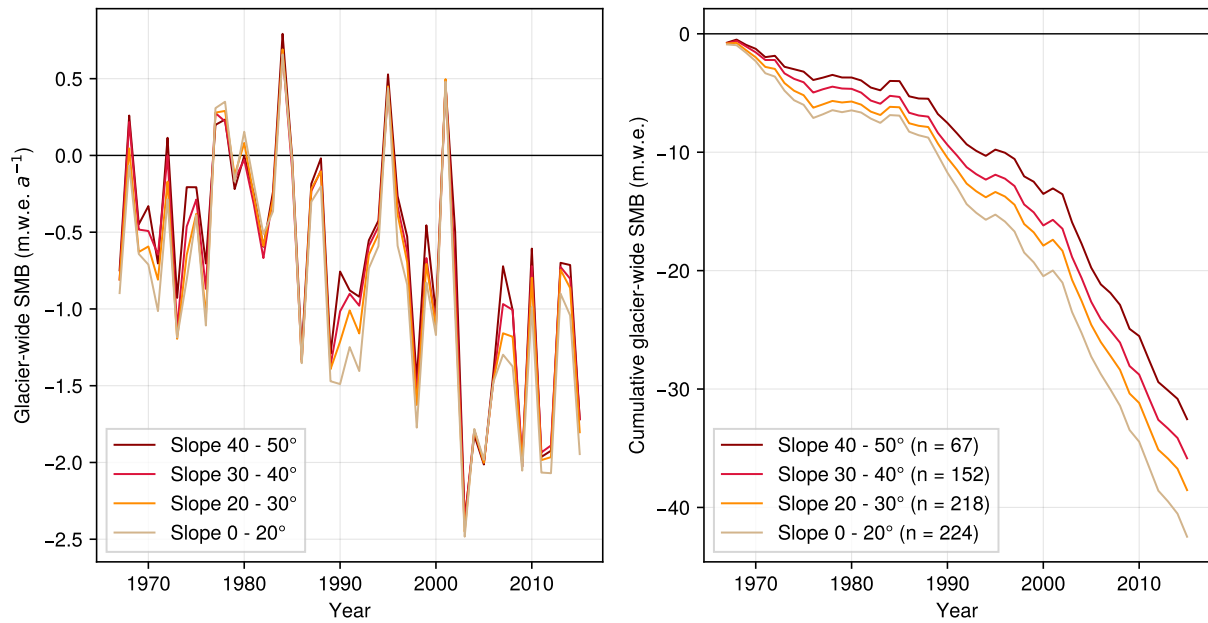
Figure S9. Average annual glacier-wide SMB for classes of glacier slope of the lowermost 20% altitudinal range (i.e. a proxy of the glacier's tongue slope)
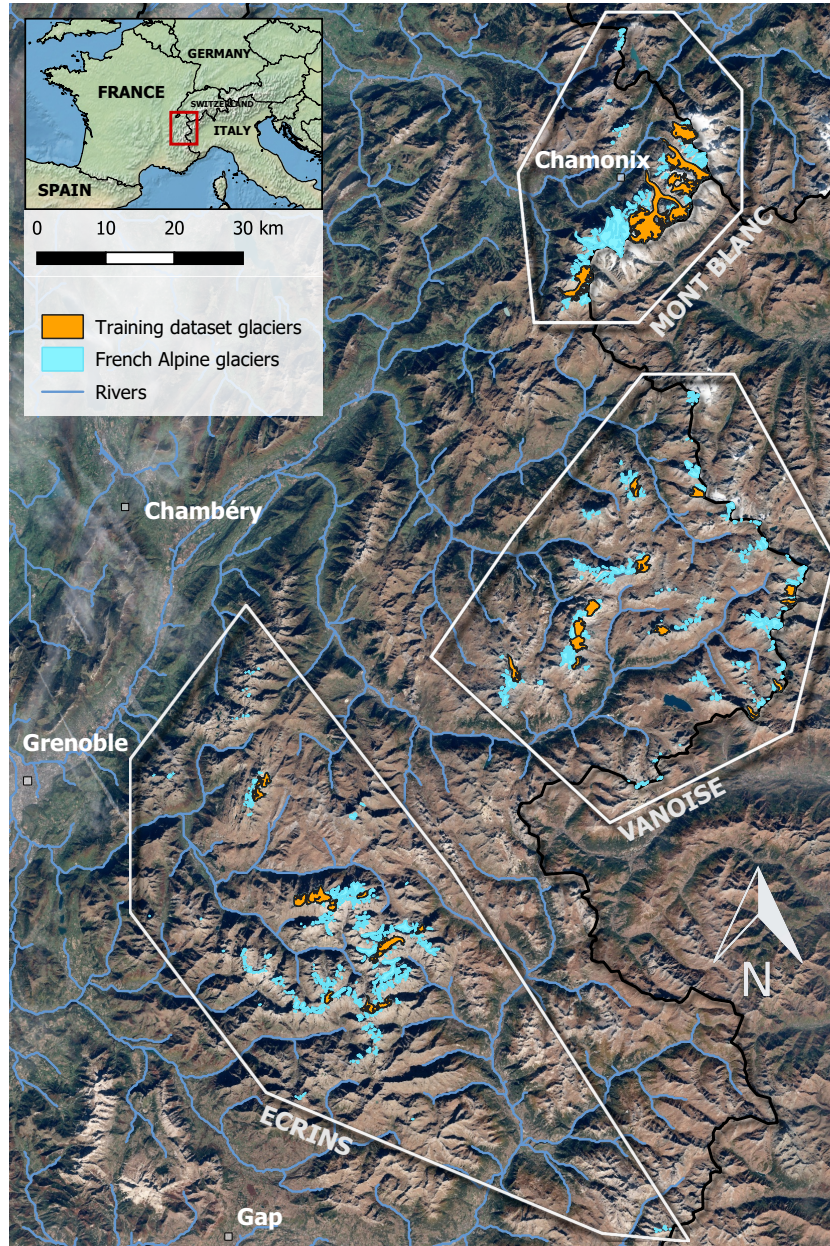
Figure S10. French Alpine glaciers used for model training and validation and their classification into three clusters or regions (Écrins, Vanoise, Mont-Blanc). Coordinates of bottom left map corner: 44º32' N, 5º40' E. Coordinates of the top right map corner: 46º08' N, 7º17' E.

References

Berthier, E., Vincent, C., Magnússon, E., Gunnlaugsson, , Pitte, P., Le Meur, E., Masiokas, M., Ruiz, L., Pálsson, F., Belart, J. M. C., and
Wagnon, P.: Glacier topography and elevation changes derived from Pléiades sub-meter stereo images, The Cryosphere, 8, 2275–2291,
https://doi.org/10.5194/tc-8-2275-2014, https://www.the-cryosphere.net/8/2275/2014/, 2014.

Bolibar, J., Rabatel, A., Gouttevin, I., Galiez, C., Condom, T., and Sauquet, E.: Deep learning applied to glacier evolution modelling, The
Cryosphere, 14, 565–584, https://doi.org/10.5194/tc-14-565-2020, https://www.the-cryosphere.net/14/565/2020/, 2020.

Davaze, L., Rabatel, A., Dufour, A., Arnaud, Y., and Hugonnet, R.: Region-wide annual glacier surface mass balance for the European Alps
from 2000 to 2016, Frontiers in Earth Science, 8, 149, publisher: Frontiers, 2020.

Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B., and Mérindol, L.: Reanalysis of 44 Yr of Climate in the French
Alps (1958–2002): Methodology, Model Validation, Climatology, and Trends for Air Temperature and Precipitation, Journal of Ap-
plied Meteorology and Climatology, 48, 429–449, https://doi.org/10.1175/2008JAMC1808.1, http://journals.ametsoc.org/doi/abs/10.1175/
2008JAMC1808.1, 2009.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset:
UPDATED HIGH-RESOLUTION GRIDS OF MONTHLY CLIMATIC OBSERVATIONS, International Journal of Climatology, 34, 623–
642, https://doi.org/10.1002/joc.3711, http://doi.wiley.com/10.1002/joc.3711, 2014.