Reviewer 2 – Fabien Maussion

[RC2.01] The manuscript by Zekollari and colleagues presents new estimations of projected glacier change in the European Alps. It is a well conducted study, the paper is well written and the results are interesting. The inclusion of ice dynamics, the use of CORDEX data instead of coarser GCMs and the large amount of calibration data are the main novel points in this study. It will become the new reference study for future glacier change in the European Alps and as such, it is likely to receive a larger interest from the general public and the media. I have two major concerns that need to be addressed before publication, as well a several specific questions / recommendations.
We thank the reviewer for his generally positive appreciation of the paper. We have addressed the two major points as well as all other specific comments in the revised version of the manuscript.

General comments - Validation and uncertainty

I acknowledge the efforts realized to use so many different observational datasets, an exercise only possible in the European Alps. However, I have several issues with the model validation in this study: **[RC2.02]** 1. there is no information about how many glaciers (and how much ice area/volume) are simulated without any calibration data. For these glaciers, the geodetic MB is "interpolated" and the effect of this interpolation is not assessed
In total, 1508 glaciers have a geodetic mass balance that can be used for SMB model calibration. By number, this corresponds to 38% of all glaciers. The distribution of glaciers that have a geodetic mass balance, however, is skewed towards larger glaciers, and as a consequence, 60% of the total Alpine glacier area has a geodetic mass balance. We have added this information in the updated manuscript: *About 1500 glaciers (ca. 38% by number) have a glacier-specific geodetic mass balance observation. Since larger glaciers are overrepresented in this sample, however, this corresponds to about 60% of the total Alpine glacier area.* What concerns the interpolation of the geodetic mass balance to glaciers without direct observations, the effect is now explicitly addressed (cf. RC1.04). In the updated manuscript, the following passage was added: *When only considering SMB measurements on glaciers that have no observed geodetic mass balance (i.e. glaciers for which the geodetic mass balance used to calibrate the model was extrapolated from other, nearby glaciers), the misfit between modelled and observed values increases only little (RMSE = 0.79 m w.e. yr^{-1}; MAD= 0.72 m w.e. yr^{-1}; mean misfit = -0.19 w.e. yr^{-1}), indicating that the method used to extrapolate the geodetic mass balances to unmeasured glaciers performs well.*

[RC2.03] 2. there is no indication as to the computation of the uncertainty ranges provided in the glacier changes (e.g. in the abstract). Does it originate from the forcing ensemble? The model RMSE? It is hard make further assessments without this information
It is true that the nature of the uncertainty ranges was not clearly formulated. These uncertainties result from the ensemble of RCM simulations, and this is now explicitly mentioned in the abstract: *We find that under RCP2.6, the ice loss in the second part of the 21st century is relatively limited and that about one-third (36.8% ± 11.1%, multi-model mean ± 1σ) of the…* And in the text (when the results are presented for the first time, in section 5): *Under RCP2.6, in 2100 about 65% of the present-day (2017) volume and area are lost (–63.2±11.1% and –62.1±8.4% respectively, multi-model mean ± 1σ,…*

[RC2.04] 3. the validation using observed traditional MB does not make sense to me, because the model has been calibrated on geodetic MB on the same glacier (both data are not exactly the same, but close - see also the comment of Ben Marzeion). If anything, you should use cross-validation here: when assessing the performance of the model on a given glacier, you remove the selected glacier from the calibration dataset, then use this data plus the traditional MB measurements to assess the model. This would also help to address point 1
We agree that the SMB calibration was not fully independent, and have therefore changed the calibration procedure as suggested by the first reviewer (cf. RC1.04). We now make a distinction between different types of validation data, and for validation we only consider SMB observations that do not temporally overlap with geodetic mass balance measurements. Additionally, a comparison between the observed and modelled SMBs is also made for glaciers without any geodetic mass balance observation, which shows that the extrapolation method for geodetic mass balance works well. For more details, we refer to our replies to RC1.04 and RC2.02, where the textual changes are also described.

[RC2.05] 4. it is problematic that the effect of the RCM forcing is not assessed at all. The plots all start in 2017, so any sceptic reader could say: "this is all extrapolated without test in the past". I understand the problems behind the validation of RCM forcing because of internal variability, but: since you are bias correcting over a reference period, at least the MB model bias (not RMSE) could be assessed when driven by RCMs as well for glaciers with long observation time series. These data would provide a much better estimate of the true uncertainty of the model driven by RCM data for the future. I'm leaving it open to the authors if they want to implement this validation or not - I believe it would make their paper much stronger.
The idea of adding an analysis of the RCM data in the past is an interesting one. We have incorporated this in part: Rather than performing such an analysis through an SMB validation procedure, we have added a comparison between: • Past Alpine-wide SMBs obtained by forcing the SMB model with E-OBS data • Past Alpine-wide SMBs obtained by forcing the SMB model with RCM data For the latter, we decided to use the historical runs of the EURO-CORDEX models (instead of forcing with ERA-INTERIM). This ensures that the RCM-model skill is assessed, rather than the quality of ERA-INTERIM, and allows for a long comparison period. The comparison shows that, the general tendency and interannual spread in SMB obtained when forcing the SMB model with historical RCM simulations, is comparable to the one when forcing the SMB model with E-OBS data. This is now described in the manuscript, and a figure was added to the supplementary material (suppl. mat. Fig. S2 in the updated manuscript): *Finally, sensitivity tests were performed with the SMB model being forced with historical RCM output (instead of E-OBS). The tests indicate that the RCMs, despite not being forced with reanalysis data, are producing general SMB tendencies that are relatively close to those obtained when forcing the model with E-OBS data (similar mean values, see suppl. mat. Fig. S2; similar interannual variability: $\sigma_{SMB,EOBS}$ = 0.66 m w.e. yr^{-1}*; mean $\sigma_{SMB,RCM}$= *0.58 m w.e. yr^{-1}).*

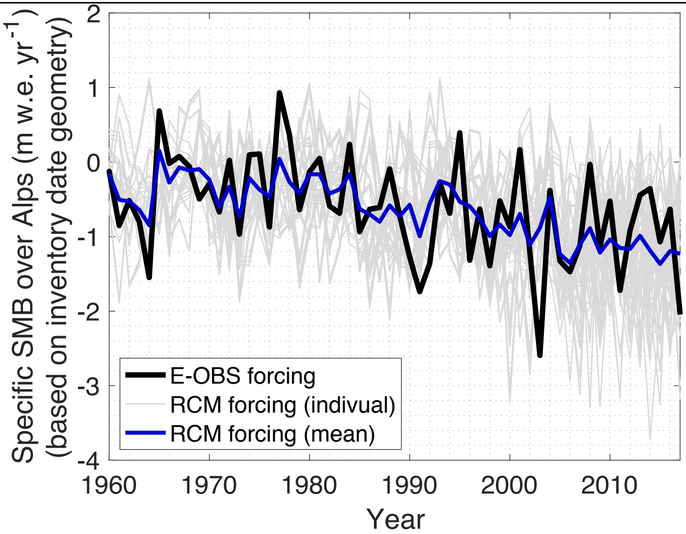

Fig. S2. Past specific surface mass balance (SMB) over glaciers in the European Alps, as derived from E-OBS data and from historical RCM simulations. SMB calculations are based on reference geometry at inventory date.

General comments – Glacier geometry

[RC2.05] The Huss and Farinotti (2012) approach (HF2012), which is to "squeeze" glaciers into elevation bands is an interesting compromise parametrization, simpler than the multiple flowline algorithm followed by OGGM (Maussion et al., 2018) but still allowing for ice flow considerations. It has some advantages (I don't necessarily agree with the ones listed in the paper): it is programmatically more efficient, arguably more elegant (because simple), and it is probably less sensitive to uncertainties in glacier outlines or topography. It also has some disadvantages (mostly, the lost of geometrical information for more complex MB models, and the over simplification of the mass flow along multiple branches).

In an attempt to reproduce the method following the algorithm description by HF2012, I consistently obtain shorter glaciers than provided by the authors (e.g. as shown in Fig. 5). See https://nbviewer.jupyter.org/github/fmaussion/misc/blob/master/simplified_flowline_tests.ipynb for some code and graphics.

I wonder why I can't reproduce the authors' results, and I therefore have a few questions:

- what motivated the choice of 10m for the δz elevation bands? This is quite a narrow range and I get better results with larger bands (depending on the underlying map resolution)

The reviewer points out some interesting differences between the flowline approach used by OGGM and the one we use. We acknowledge that some of these points were not included in our original submission, and now do so in the reformulated text:

Subsequently, the glacier geometry is interpolated to a regular, horizontal grid along flow. Through this approach, possible glacier branches and tributaries are not explicitly accounted for, avoiding complications and potential problems related to solving the little-known mass transfer in these connections. As such, this approach is less sensitive to uncertainties in glacier outlines and topography compared to methods in which glacier branches are explicitly accounted for (e.g. Maussion et al., 2018), but may in some cases oversimplify the mass flow for complex glacier geometries (e.g. with several branches).

In this and the following answers, we argue why the reviewer may have obtained other glacier lengths compared to us, and explain how we updated the manuscript to further clarify the H&F method applied:

- For the difference in obtained glacier lengths, we refer to our response to RC2.07
- The 10-m elevation bands were chosen to ensure that the method is consistently applicable, also for small glaciers.

- **[RC2.06]** what do you do when there is no glacier grid point in a 10m band? This happens quite often depending on the underlying map resolution (see graphs).

Having 10-m elevation bands without a glacier grid is not a problem in our case, as the geometric representation with elevation is transformed to a grid with a constant horizontal spacing. This is mentioned in the manuscript:

The horizontal distance (Δx) between the elevation bands is determined from the elevation difference (Δy) and the local surface slope (s):

$\Delta x=\Delta y/\tan s$. (1)

Subsequently, the glacier geometry is interpolated to a regular, horizontal grid along flow.

In case values are missing (i.e. no area and volume for a particular elevation band), these are simply neglected during this interpolation procedure.

- **[RC2.07]** do you do any kind of filtering for large slopes? The skewed slope distribution towards high slopes can affect the mean and, together with the missing bands, could explain why I get shorter glaciers.

A filtering of the local slopes is performed to get the average slope of elevation bands (that are subsequently used to compute glacier length). This filtering was indeed not described in detail in the original publication (Huss and Farinotti, 2012), thus hampering complete reproducibility. To determine band-average slope, all values below the 5% quantile are discarded, as well as all values above a threshold (typically around the 80 to 90% quantile) determined based on the skewness of the slope distribution function. The approach reduces the effect of very steep cells within an elevation band on average band slope and, hence, glacier length, and has been optimized based on comparisons to flowline glacier length.

This is now formulated after Eq. (1):

To determine the band-average slope s, all values below the 5% quantile are discarded, as well as all values above a threshold (typically around the 80 to 90% quantile) determined based on the skewness of the slope distribution function.

- **[RC2.08]** do you do apply any smoothing on the resulting band widths and areas? They appear quite noisy in my case (depending on the underlying map resolution).

No smoothing is applied of the glacier bands and widths. Despite the fact that the band widths and areas can strongly vary in space (being 'noisy'), they do not lead to any numerical problems when solving the transport equation.

[RC2.09] I'd like to see these questions answered in this manuscript, unless I missed them from either HF2012 or Huss and Hock (2015), in which case I'm happy to be corrected and pointed to the location where the algorithm is described.

Similarly, there are some locations in the current manuscript where I find that the algorithm description is too vague to allow reproducibility (see specific comments below).

By having addressed the comments formulated above and having updated the manuscript accordingly, we hope that the reader will understand the various steps. All specific comments formulated below have been addressed.

Specific comments

[RC2.10] Abstract L10 : "which the latter" sounds strange. Rephrase?
This has been reformulated to: *…ice flow processes, of which the latter is to date…*

[RC2.11] Abstract L20 "RCM that is coupled to it" → the RCM is not "coupled" to the GCM (this suggests two-way nesting) ; maybe "nested in", or "driven by" the GCMs? Also revise other occurences in the text.
We have now changed this by omitting the 'coupling' part: *…determined by the driving global climate model (GCM), rather than by the RCM, and…* This was also updated for other occurrences in the manuscript: *… an RCM driven by a GCM…* (second last paragraph of introduction) *… on the driving GCM than the RCM, and …* (last sentence of section 6.1) *…driving GCM (rather than the RCM), and…* (conclusion)

[RC2.12] P2L13 "the evolution of the glacier" → "glacier evolution"
Modified as suggested.

[RC2.13] P2L21 "moderate" and "extreme" are subjective adjectives → be more precise, e.g. RCP or similar
This was now modified to: *These regional and global studies generally suggest a glacier volume loss of about 65-80% between the early 21st century and 2100 under a moderate warming (RCP2.6 and RCP4.5), and an almost complete disappearance of glaciers under warmer conditions (RCP8.5).*

[RC2.14] Legend Fig 1 updated version OF Huss and ...
Modified as suggested.

[RC2.15] P3L23 "we aim at reducing the considerable uncertainties" → I'm yet to be convinced that increased complexity reduces uncertainty, and I'm not sure your study really deals with this topic or even actually shows that uncertainties are reduced. It's okay if you leave this sentence as is, but you don't need this paragraph to justify your study
We understand the point raised by the reviewer, and now state that we aim at improving future projections and at examining how this could affect global glacier projections (which is indeed more what we do instead of 'reducing the uncertainties'): *Through novel approaches in terms of (i) climate forcing, (ii) inclusion of ice dynamics, (iii) the use of glacier-specific geodetic mass balance estimates for model calibration, and by (iv) relying on a vast and diverse dataset on ground-truth data for model calibration and validation, we aim at improving future glacier change projections in the European Alps. As a part of our analysis, we explore how the new methods and data utilized could affect other regional and global glacier evolution studies.*

[RC2.16] P4L3 what is the "local surface slope"? According to HF2012 it is the bin average

for each elevation band. Be more precise in the formulation here (see also general comment about that).
This comment has been addressed in our reply to RC2.07. In particular, the manuscript was updated in order to clarify how the local surface slope is determined.

[RC2.17] P4L7 "little-known connections" → I don't understand what you mean. Connections are maybe more complex in a dynamical sense but so are other locations on the glacier as well. Furthermore, "ignoring" these connections is not making them less complex, it's just avoiding them. So, I suggest to remove this sentence (see also general comment)
The first reviewer also pointed this out, and we agree that the formulation was not very precise. It is not the connections that are little-known, but rather the mass transfer in these zones. The sentence now reads: *…complications and potential problems related to solving little-known mass transfer in these connections.*

[RC2.18] P4L9 trapezoidal sections: how does this go together with the ice thickness inversion? What cross-sections are used in HF2012? If rectangular (I assume), by using a trapeze you are either reducing the sections volume or increasing the thickness h0, i.e. you are not physically consistent between the inversion and the forward model.
This is a valuable comment, and we agree that it was not clearly formulated how we treat the different cross section parameterizations. In Huss and Farinotti (2012), a rectangular cross section is used, while we rely on various cross section representations (trapezium with different shapes and rectangular cross section to test for sensitivity to this). In all cases, the cross-section transformation is performed by preserving the area and volume for the particular location. This can lead to slightly different bedrock elevations in this representation, although the differences are in general minor. This is now clarified in the updated manuscript: *Glacier cross-sections are represented as symmetrical trapezoids. The bedrock elevation is determined in order to ensure local volume and area conservation.*

[RC2.19] P4L17 "close representation of past temperature and precipitation and certain events" → Reanalysis datasets also represent weather events well thanks to data assimilation. It's okay to use ENSEMBLES, but you should argue otherwise, maybe because of uncertainties in quantitative precipitation estimates or the coarse resolution of reanalysis data, for example.
In line with the reviewer's suggestion, this has now been reformulated to: *This E-OBS product represents past events closely (for example the heat wave of the summer of 2003, Fig. 2b), allowing for detailed comparisons between observed and modelled surface mass balances (section 4.1). We prefer using an observational dataset compared to a re-analysis product (e.g. ERA-INTERIM, as used in Huss and Hock, 2015), as the former has a higher resolution and goes back further in time.*

[RC2.20] P4L24 I think this whole justification paragraph is more confusing than helping. I think it's okay to use an observational dataset for calibration and validation instead of reanalysis, consider shortening this paragraph.
The paragraph was shortened when addressing the reviewer's previous comment (see RC2.19).

[RC2.21] P4L28 is "chains" the commonly used word for this? I thought that "realisation" or

"simulations" would be more appropriate. See also other occurrences in text.

This was now modified to '*simulations*' throughout the text. The wording is classically used in the literature (e.g. Kotlarski et al., 2014, GMD). See also our replies to RC1.16 and RC1.17.

[RC2.22] P5 Eq. (2) I have several questions here. First, you don't say over which observational period you compute the averages for the monthly bias correction. Is it 1961-1990? The entire observation period? I assume that σobs and σobs is computed for the same reference period as the bias. Then, why choosing a 25-yr period, and not a period of the same length as the reference period?

Please also add a sentence as to why you don't apply such a correction for precipitation. I understand that the arithmetics are not so easy for multiplicative bias corrections, but in theory some kind of correction would also be possible (and might be needed by looking at Fig. 02).

The bias was evaluated over the longest possible period where both RCM data and E-OBS are available. Some RCMs are available from the 1950s on, while others only start in 1970. The period considered for computing the averages for the monthly bias correction thus ranges from 1970 to 2017, as we now explicitly mention:

This correction is applied over the period 1970-2017, which is the overlap period for which all RCM simulations and E-OBS data are available.

A similar correction is not applied for precipitation, as this is a "cumulative" quantity: i.e. monthly differences in variability will not be that relevant at the annual scale (mass budget). Furthermore, variabilities in precipitation do not have a direct effect on the calibrated parameters (as is the case for temperatures via the degree-day factors). This has now been formulated as:

For precipitation, which enters the SMB calculations as a cumulative quantity, no correction for interannual variability is applied, as the monthly differences in variability are not that relevant at the annual scale. Furthermore, variability in precipitation does not have a direct effect on the calibrated SMB parameters (as is the case for temperatures via the degree-day factors, see section 3.1.).

[RC2.23] P5L26 "based on a combined criterion weighting both horizontal distance and the difference in area." Can you be more specific here? (reproducible science versus "black box"). How many glaciers have Geodetic and traditional MB observations? Which area does it represent? How many glaciers needed this kind of interpolation?

As we stated in our responses to RC1.04 and RC2.02, we now provide more information about the number of glaciers (and the area) that have geodetic mass balance observations. We also added a more elaboration explanation about the procedure that is used to derive the geodetic MB for glaciers without such observations:

In case no geodetic mass balance observation for the specific glacier is available, an observation from a nearby glacier is chosen. The respective observation is selected based on the two criteria horizontal distance (in km) and relative difference in area (unitless). We multiply the two criteria and consider the minimum as the most suitable glacier to supply a mass balance observation for the unmeasured glacier. The replacement thus represents a nearby glacier that is relatively similar in size. The effect of this approach is evaluated in section 4.1

[RC2.24] P7L25 what kind of numerical solver are you using? It's not an harmless choice, as shown by Jarosch et al 2013.

We agree that it is important to consider stability and mass conservation issues, which are strongly related to the type of numerical solver. Implicit methods allow for using larger time

steps, while explicit methods are intrinsically less stable and need smaller time steps. The latter are however computationally less demanding, and therefore more efficient (see e.g. Schäfer et al., 2007, JGlac). We use a semi-implicit solver. For the calculation of the continuity equation (eq. 6), it relies on an intermediate time step during which the geometry is adapted. We now explicitly mention this in the manuscript.
The continuity equation is solved using a semi-implicit forward scheme by relying on an intermediate time step (i.e. sub-time step update) in which the geometry is updated.

[RC2.25] P8L10 "Notice that through this approach, the glacier is not assumed to be in steady state at any point in time, but that an artificially modelled steady state is obtained by imposing a MB offset." → I don't understand what you want to say here. I'm also quite confused at the statement "A determines volumes, SMB bias determines the length". Is this based on you own experience, or is there a physical explanation? Finally (and most importantly), why is length used as convergence criterion instead of area, which is the only variable which is almost perfectly known at the inventory date?

With the sentence formulated on P8L10 in the original manuscript, we want to stress that the steady state that we produce is an artificial one, which we impose for our calibration procedure (to match the geometry at inventory date, see also our response to RC1.03), but we do not assume that the glacier was in steady state with any climatic conditions. In hindsight, we agree that the original sentence was not very clear, and we have decided to omit it altogether.

The second statement ("[parameter] A determines volumes, [the] SMB bias determines the length"), is based on physics, where the rate factor (A) determines the stiffness of the ice, and through this the local ice thickness and thus the ice volume. This is also evident from the equations: if A increases, the local velocity or flux increases, resulting in lower ice thickness. By modifying the SMB bias to create the artificial steady state, the length of the steady state is modified (this is because the SMB needs to be zero when integrated over the glacier). A different steady-state length, in turn, causes the length of the modelled glacier to be modified at inventory as well. We have now clarified this as follows:
The glacier volume and length at inventory date are matched by calibrating two variables (Fig. 3). The first calibration variable is the deformation-sliding factor A, which mainly determines the volume of the glacier at the inventory date. The reason for this resides in the role that A has on the local velocity/flux, which in turn affects the local ice thickness and thus the ice volume; see Eqs. (4-7). The second calibration variable is an SMB offset in the 1961-1990 climatic conditions used to construct a 1990 steady-state glacier, which mainly determines the length of the steady-state glacier (as the geometry is such that the integrated SMB equals zero). Note that a change in steady-state length causes the glacier length to change at inventory date as well.

For the comment related to the use of lengths (vs. areas) for calibration, refer to RC1.03.

[RC2.26] Model initialisation needless to say, the iterative initialisation procedure is... unconventional. I'm not asking to change it, because it serves one purpose: find a transient glacier which is consistent with the forward model at a reference date. This is necessary because the ice-thickness inversion model and the forward model in GloGEMFlow are probably not consistent between each other (different MB profiles, different A, different bed shapes).

However, I would like to add that I don't really think that this iterative method has much to do with finding an "appropriate" A for each glacier. Let's take the first step as an example: since you drive your model with an SMB such that the present day geometry is in equilibrium, modifying A so that your glacier has to grow will always tend towards lower values of A in order to create a thicker, longer equilibrium glacier in 1990.

The reviewer suggests that by modifying *A*, the glacier length will be modified. This is rarely the case, as *A* mainly determines the glacier thickness. Through this, it can slightly influence the glacier length (through the SMB – elevation feedback), but this effect is much smaller than the effect that the SMB bias has. We have now clarified this by adapting our manuscript, as we explain in our previous response (reply to RC2.25).

[RC2.27] P8L23 this cannot be considered an "independent" validation (see general comment)
See RC1.04: We agree and have now reworked the evaluation procedure by discarding observations that temporally overlap with geodetic mass balances

[RC2.28] P8L25 "rather than the coupled SMB – ice flow model" → this is a bit of a missed opportunity, because there are chances that the varying geometry actually improves the SMB validation, by taking geometry changes into account which are present in observations but not in the static model.
See RC1.25: There are several reasons for which we decide not to rely on a dynamically evolving glacier geometry for the SMB validation, and these are now mentioned in the text.

[RC2.29] Fig 4 Legend r2 is the "coefficient of determination"
This was modified.

[RC2.30] P8L29 elevation bands and correlation → I agree with Ben Marzeion
See RC1.26.

[RC2.31] Fig 5 intuitively, I would swap the glacier flow direction so that the distance on model grid (x-axis) is starting from zero at the glacier top. This would also allow to read the length of the glacier directly on the x-axis
Swapping the glacier direction may indeed be an option. But as the figures does not start at zero (it rather shows the distance along the model grid, to ensure a consistency with the figures that will be added as supplementary material), we decided to leave this as it was.

[RC2.32] P9L21 how did you compute the surface velocity out of the depth-integrated velocity given by the shallow-ice approximation?
As basal sliding is not treated explicitly in our approach, and given that we assume that the mass transport is defined by the local geometry (SIA), the surface velocities ($\bar{u}$) are equal to the 1.25 x depth-integrated velocities (u_s) ($\bar{u}/u_s$ = 0.8) (see e.g. Cuffey and Paterson, 2010, p.310). This is now specified: *In the lower parts, where many glaciers have a distinct tongue, a comparison between observed and modelled surface velocities is possible (surface velocities correspond to 1.25 times the depth-integrated velocities, since we treat basal sliding implicitly, see e.g. Cuffey and Paterson (2010, p.310)).*

[RC2.33] P10L12 note that other length records are also available for the non-swiss glaciers (WGMS or Leclerq database)
We now mention the existence of other datasets: *Note that other length records are also available for non-Swiss glaciers (e.g. Leclercq et al., 2014), but that these were not considered to ensure a consistency in derived length*

records.

[RC2.34] P10L20 remove "highly significant" and the p-value to read "the correlation is r2 = 0.37 (p < 1e-3)"
This was modified as suggested.

[RC2.35] P11L4 unit km2 yr-1
Indeed! This was modified.

[RC2.36] P12L6 "highest correlation with the maximum glacier elevation" → is this sentence correct?
This should have read *highest correlation with the glacier elevation range* (cf. Table S3). We thank the reviewer for spotting this!

[RC2.37] Fig 9 Legend remove the "two" in "two present day"?
This was modified.

[RC2.38] Fig. S3 Consider adding Fig. S3 to the main manuscript.
Fig. S3 from the original manuscript was added to the main text and is now Fig. 10. The figure numbering the main text and in the supplementary material has been updated accordingly.

[RC2.39] P14L10 what do you mean with "ice is more pronounced"?
Indeed, this sentence was not clear, as we referred to *ice* as being *more pronounced*, while it should have read *ice loss...more pronounced. We* now modified this accordingly.

[RC2.40] P15L12 when the variable IS considered? I'm not sure I fully understood this section.
We thank the reviewer for pointing this out. This should indeed be 'IS considered' (vs. is <u>not</u> considered in the original manuscript). The text now reads: *In such an analysis, all independent variables are replaced by dummy variables, which have a value of one when the variable is considered, ….*