

Reviewer 1

General comment

[RC1.01] This is an interesting and well-designed paper by Zekollari et al, which provides a comparison of global glacier volume change under CMIP6. Additionally, the paper investigates the use of glacier-specific calibration techniques. Overall, the paper well written with the results thoroughly discussed. I have no major comments, but think the clarity of the text could be improved in certain places and have listed some minor comments below (including some adjustments to figures).

We thank the reviewer for their positive comment and general appreciation of the manuscript. We will improve the clarity of the manuscript by following the provided suggestions.

Specific and technical comments

[RC1.02] L90 – I find the description of this methodology a bit confusing, please consider rephrasing

This sentence will be rephrased to:

The advantage of this approach is that the obtained mass balance parameters have (physically) realistic values, which fall within the literature ranges

[RC1.03] L136 – Please state the year which the ice thickness dataset represents

The ice thickness is at the Randolph Glacier Inventory (RGI) date, which varies among (and also within) regions. This information will be added in the updated manuscript:

In GloGEM, the ice thickness is from the consensus estimate of Farinotti et al. (2019a) at the RGI inventory date, which is deduced from the surface elevation (as provided in Farinotti et al., 2019a) to reconstruct the bedrock elevation

[RC1.04] L168 – Please specify when the past climate starts (e.g. XXXX until 2020)

This information will be added:

...(1980 until 2020)...

[RC1.05] L171 – Please specify when the future climate ends

This will be included in the updated manuscript:

(from 2020 until 2100)

[RC1.06] L180 – What is the debiasing procedure?

The debiasing procedure allows for a consistency between the past climate data (ERA5) and the future climate model data over the common time period. This consistency occurs through a procedure that is described in the original GloGEM study (Huss and Hock, 2015). This will be further clarified in the updated manuscript:

To ensure consistency between the observational/past (ERA5) and the future climate model data, a debiasing procedure is applied over the common 2000-2019 time period following the procedure described in Huss and Hock (2015)

[RC1.07] L192 – I think ‘Alternatively’ should be changed to ‘Additionally’

This should indeed be ‘Additionally’, which we will change in the updating the manuscript:

Additionally, we also evaluate...

[RC1.08] L227 – The sentence beginning ‘In general, if...’ is phrased in a slightly confusing way and could be reworked for better clarity – maybe something like ‘In general, the mass balance calibration parameters are calibrated to give a more negative SMB for glaciers which have a lower mass balance than the regional average. This translates...’

This sentence was indeed a bit confusing. We will reformulate it along the lines you suggested, but will keep it as a single sentence to allow to make the contrast between this case (mass balance glacier < regional one) and the opposite case (mass balance glacier >

regional one, which we refer to as *and vice versa for higher mass balance*). The update sentence will be:

In general, for glaciers with a mass balance lower than the regional one, the mass balance model parameters are calibrated to produce a more negative present-day mass balance, which translates into a more negative future mass balance and thus more substantial projected ice loss, and vice versa for higher mass balance

[RC1.09] L284 – I would change the sentence 'For most other regions, these differences are even more outspoken..'. I do not think outspoken is a good choice of word, and this sentence is a bit dense and confusing to read. Could be changed to something like 'For most other regions, there is an even larger proportion of glaciers which show differences of more than 10% in their 2015-2050 volume projections. For instance, in High mountain Asia...'

We thank the reviewer for this suggestion and will change the text along these lines to:

For most other regions, there is an even larger proportion of glaciers which show large differences in volume projections. For instance, in High-Mountain Asia (RGI regions 13, 14, 15), between 35-55% of all glaciers (with volume >0.1km³) have differences in the 2015-2050 volume projections of more than 10% depending on the calibration approach (Table S 3).

[RC1.10] L288 – This could also be rephrased/ broken up into two sentences. Else, add '..independent of calibration methodology' or similar.

This will be updated following the second suggestion to:

When considering the 2015-2100 volume evolution, the differences resulting from the calibration approaches are generally smaller, since a lot of the regions lose a large part of their mass, evolving to a similar (almost ice-free) state independent of the calibration methodology (Table S 3)

[RC1.11] L299 – I do not understand what you mean by 'since discharge is calculated over initial glacier area'

The discharge values are calculated over the initial glacier area: i.e., even if the glacier becomes smaller (e.g., evolving from 10km² to 5km²), the discharge is calculated over the initial area (10 km² in this example, accounting for the precipitation and snow melt over the deglaciated area). In the updated manuscript this will be explained by providing the following additional information:

since discharge is calculated over initial glacier area, differences in precipitation result in differences in discharge

Additionally, we will also provide additional information at the beginning of the paragraph when introducing the methodology to calculate discharge:

...calculated over the initial glacier area, i.e. with fixed watershed area, following the method presented in Huss and Hock (2018), accounting for glacier and non-glacier runoff

[RC1.12] Fig.2 – Please make the text in the 2015-2100 volume change boxes larger

In the updated figure, we will increase the font size of the text in the 2015-2100 volume change boxes

[RC1.13] Fig. 3 – Please also make the text large in this figure (for the calibrated values)

The text for the calibrated values has been made larger.

[RC1.14] Figs 2 and 3: It could be beneficial to move some of the panels to the supplementary material, so that the remaining panels can be more easily seen – as of now, they are quite busy.

Whereas these figures contain quite a lot of information, we would like to keep the same number of panels to clearly reflect the contrasts in the effect of the calibration for different types of glaciers in the main text. By increasing the font size of the figures (see previous comments), we have increased the readability of the figure.

[RC1.15] Fig.7 – If possible, please extend the width so that it takes up the full page width/ has the same dimensions as Fig.6

We will extend the width of both figures so that they take up the full page width.

[RC1.16] L424 – The explanation of why the inclusion of frontal ablation in GloGEM leads to less global volume loss does makes sense, but I had to read it a few times to understand. It would be good to add in a bit more detail here to make it extra clear.

In the new manuscript we will update this to contain additional detail:

In GloGEM, since frontal ablation contributes to the total mass balance, a higher surface mass balance is needed to result in the same total mass balance as for the case without frontal ablation. As a consequence, if frontal ablation decreases (e.g. loss contact with ocean), the more positive mass balance dominates (vs. case without frontal ablation), resulting in less future ice loss. Given the very large uncertainties in modelled present-day and future frontal ablation, it is currently difficult to judge whether results from a setup with a relatively uncertain frontal ablation (GloGEM) or one in which it is not explicitly represented (OGGM setup used in this study) should be more trusted.

[RC1.17] L468 – Formatting error (sentence beginning with ‘Table1Figure7A’)

Thank you for spotting this. This will be corrected in the new manuscript.

[RC1.18] Fig. 9 – It would be nice to move the GloGEM panel next to the panel which compares CMIP5 GloGEM to CMIP model ensemble, for ease of comparing the GloGEM results with each other. Also include a ‘CMIP6’ label under the GloGEM results like in the other panels. It could additionally be beneficial to add results showing the mean of the CMIP6 forced GloGEM/OGGM/PyGEM simulations, to be compared to the CMIP5 ensemble mean. In this case, you could move the CMIP5 GloGEM results to the GloGEM panel and then have an ‘ensemble mean/ CMIP5 vs CMIP6’ panel.

L519 – Here you compare CMIP5 ensemble to your new CMIP6 results, which I think would be strengthened through the above changes to Fig. 9. I understand the caveats to direct comparison that you explain in the text, but still think this would be a useful visual aid.

Adding GloGEM CMIP5 simulations to Figure 9 would be interesting to allow for a direct comparison on the effect of the CMIP5 vs. CMIP6 forcing for the GloGEM results. However, we do unfortunately not have the data at hand for doing so. More specifically, we did not re-run the latest version of GloGEM (which, among other novelties, now also includes a glacier-specific calibration) with CMIP5 simulations: the CMIP5 GloGEM simulations that we have at hand are those that we submitted at the time for GlacierMIP2, i.e. relying on an older version of the model (and without glacier-specific data at hand for calibration). As such, moving these (old) CMIP5 simulations (now the right-most column of the figure) to the GloGEM part of the figure would be misleading, as comparing this to the new CMIP6 simulations would be “comparing apple and oranges” (differences is volume can result from differences in climate forcing, model differences, and/or differences in model calibration).

For the two other models (OGGM and PyGEM), the model was re-run with exactly the same model version and calibration data for CMIP5 and CMIP6, therefore allowing for a comparison (and hence they are displayed next to one another). For the above reasons, we cannot aggregate the different model versions to create an ensemble mean for CMIP5 vs. CMIP6. We agree that this would have been an interesting comparison (if CMIP5 GloGEM simulations with latest version would have been available), but would like here to stick to the current organisation for the sake of clarity and to allow for ‘clean’ comparisons. We also believe that the current figure version accurately displays which data comes from which study: GloGEM (this study), OGGM v1.6.1 (this study), PyGEM (Rounce et al., 2023), and GlacierMIP2 (Marzeion et al., 2020; multi-model ensemble mean + global results from GloGEM version that was available at the time of that study). Following your suggestion, we will add the fact that GloGEM in this study is CMIP6 (i.e. add label below the bullet points), and also correct a typo for the CMIP5 GloGEM simulations in GlacierMIP2.

[RC1.19] L600 – This is a very long sentence that could benefit from being broken up into a few parts

This sentence will be split in two parts:

In this respect, new glacier outlines (RGI v7.0, RGI Consortium, 2023) and other datasets with a (near) global coverage (e.g., on ice surface velocities and ice thickness reconstructions (e.g. Millan et al., 2022), and frontal ablation estimates (Kochtitzky et al., 2022)) will allow inverting glacier properties, calibrating model parameters, and evaluating model performance in a more advanced way. To combine this broad variety of datasets and observations, the field of large-scale glacier modelling will increasingly rely on machine learning and data assimilation (e.g. Bolibar et al., 2023; Cook et al., 2023; Jouvét and Cordonnier, 2023).

[RC1.20] Fig. S3 – Make this figure take up the full page width; the panels are too small at the moment

This will be updated in the new manuscript.