

Malles & Marzeion, Global glacier mass change, TC

This paper provides a global 20th century glacier mass balance reconstruction driven by an ensemble of climate data sets. Few studies exist that estimate the time series of 20th century global-scale mass changes. It is an important and excellent paper that should be published after some revisions. Overall it is very well written, however, I am not quite sure about the methodology, which is not clear to me (see below).

General

1. **Methodology/structure:** The methodology is somewhat confusing in particular calibration and validation. 2.1.2 is about mass balance calibration, then 2.2. is about data and then 2.3 about optimization which I assume is 'calibration'. However the 3. Paragraph in 2.3. talks about validation. All this is confusing. I suggest to avoid that 2.2 Data is sandwiched between calibration and validation and that everything about calibration/optimization is in the same section. I suggest to have 'data' before model description/calibration. It was difficult to follow the description of the model calibration not knowing what data are actually used.
2. **Data:** it is unclear what mass-balance data are used. Are in situ measurements just those derived by the glaciological method, i.e. ignoring all the thousands of geodetic balances? The WGMS is referenced which includes geodetic balances. Also what is the temporal resolution. Are annual balances compared? Seasonal?
Overall, far too little information is given about the mass balance data. How many glaciers? How many annual measurements? Some info on the temporal distribution? How much is e.g. before 1950 or 1930. All this information is important to evaluate the methodology. Perhaps a figure can illustrate the data density somehow.
3. **Calibration:** It is unclear how the model was calibrated? Please clarify which steps are done for which glacier, which parameters are glacier-specific and what is compared? I suggest that the model is introduced first entirely independent of how parameters are obtained (section 2.1.1), then a section about calibration that includes all info how the parameters are obtained. This is currently dissected and difficult to follow. This may then allow to avoid repetition of Eq 1 and 4. It is unclear how equation 4 is applied. Where do the parameters come from? Are parameters optimized (matching observations) prior to applying eq 4? Overall this section appears the weakest and hard to grasp what was done how.
4. **Sea level contribution sign convention:** It appears that glacier mass losses expressed in SLE are treated as negative numbers. This is not consistent e.g., with the mass balance glossary (Cogley et al., 2011) and all IPCC reports. When a glacier loses mass, the mass change is negative but the resulting contribution to the ocean is positive. While a loss for the glacier system, it's a gain for the ocean. All SLE figures in the current manuscript can be misinterpreted since a sea level rise in the current paper is shown with a drop. This should be adjusted in all figures and the text for consistency with the literature.

Details

1. It should be clearer what the domain is of this paper: all glaciers outside the ice sheets but including those in the periphery of Greenland. I also suggest to replace Greenland by Greenland periphery in all figures/tables to avoid confusion. Also make clear somewhere which glaciers in Greenland you include (RGI connectivity level?)
2. Abstract line 10: What is 'temporal' here? Of annual balances?
3. Intro: Line 26: glacier instead of glaciers'
4. Intro: line 27: rephrase 'more distant past' since this can be misunderstood as thousands of years or even more ago
5. Intro: line 27-28: the statement is a bit odd. If there are no insitu measurements but we had entire 20th century satellite measurements it would be fine, so the logical connection here needs some work.
6. Line 28: in-situ measurement densities → in-situ measurements
7. Line 30: mass loss → mass change
8. Line 41: remove yet
9. Intro: Line 71: better? 'five model parameters'. Unclear what is meant by global?
10. 2.2.1: Line 122: Where do initial values for volumes come from? I assume only glaciers currently existing in RGI6 are considered, i.e. glaciers that have melted are ignored?
11. Line 124: add version of RGI and reference.
12. Line 134: I assume you allow some uncertainty around 0? Important to add that this is specific mass balance (the value depends on the unit): → 'in a zero specific annual mass balance
13. Line 141: add 'annual' to mass balance
14. Line 141 and many other places: the term 'respective' seems to be used a lot in the way it is used in German but not necessarily in English. Can be deleted in most places or in some replace by 'corresponding'
15. 2.2 Data: line 178: can you give some measure how much/often
16. Lines 188-189: perhaps the 3 short sentences can be shortened/combined?
17. Line 191: annual? Glacierwide specific? See general comments. This section would benefit from considerable expansion and clarification.
18. 2.3 Optimization: Line 204: 900? How come?
19. Line 230: what is 'a respective data set'? do you mean 'validated glacier and meteorological forcing data set'?
20. 3.1. Performance: Line 249: what is 'validated meteorological data sets'. Did you validate th climate sets with insitu observations?
21. Line 252, what data sets? You mean: for most forcing data sets?
22. Line 266: a few spelling errors
23. Line 270: 55 unique pairs: unclear? Why 55?
24. Line 271: remove 'respective'
25. Line 281: what is 'low-scored model setups' ? Do you mean the meteo forcing that scored low?
26. 3.2 Difference
The title of the header is confusing

27. Line 290: 'respective' → corresponding
28. Line 329: 'for we do' ???
29. Chapter 4: Line 374: to reduce ?
30. 5 Discussion: Can the first paragraph be broken up in 2 or 3? It's long.
31. Line 387: reformulate 'more correct'. It's either correct or not
32. Line 386-387: this is not necessarily true. You seem to assume that GRACE and the observations are perfect and have no other errors
33. Lines 390ff: lakes are mentioned but the discussion appears not comprehensive. Many other processes can delay the discharge of melt water to the oceans (e.g. groundwater, dams, human uses and transfers, evaporation etc). All difficult to quantify but perhaps the discussion here can be more balanced.
34. Line 391-392: This makes it sound as if Gardner came up with the bias but did not do anything about it, only Zemp did years later. This is not correct. Because of the bias the 'best' method was used in each region and direct glaciological data only used in regions where the measurement density is high.
35. Discussion: The more negative balances compared to others in the period well-studied period 2003-2009 is striking while consistent with Marzeion et al 2012 based on the same model. This is an important discrepancy. Can this be explained in terms of the calibration process? It appears the model was validated/calibration only with direct observations? Why was the model not validated with those thousands of recent geodetic balances covering entire regions in some places.
36. Line 395-400: a number of language issues/convoluted language sometimes difficult to follow.
37. Conclusions: Line 420: what is model setup?

Figures/Tables

1. Some multiplot figures are referred to as 'upper/lower' panel, others have labels, a,b, ... Would be good to use the same method across the paper (perhaps the latter).
2. Figure 2: caption: clarify what 1 mass-balance obs is? One annual balance on one glacier? If so N is number of glaciers x number of years per glacier: would be good if that could be distinguished in the figure, or somehow visualizing N for time period classes.
3. Figure 3: can be shortened by: (a) averaged over Asia) and (b) globally
4. Figure 3: meteorological forcing data sets
5. Figure 5: I suggest to adjust the tick labels to decadal periods, i.e. not 1901 but 1900, not 1921 but 1920 Also mark year 1980 somehow (vertical line)
6. Figure 5 and 6: I suggest that the color scheme is adjusted that the median and mean stick out better (line thickness? Color choice?). They are hard to identify.
7. Figure 5: clarify in caption of the ensemble mean is with or without the outlier model (also elsewhere)
8. Figure 7: no need to shorten the legend text in both subplots. There is lots of space, so it can be fully spelled out.
9. Table 1: Resolution: what's the unit?
10. Table 2:
 - a) for respective → for each forcing data set ?

b) Last sentence is unclear (A_M). What does it mean? Able to initialize?

c) To make the table stand on its own would be good to spell out what the variables are (at least those that are not obvious) or at least group them somehow, e.g. xxxx are model parameters to determine melt and snow accumulation (Equation xxxx).

11. Table 3: different number of decimals for the same variable is unfortunate.

- caption: add? 'for 18 primary RGI regions'

12. Table 4: hard to get a grasp on trends. I suggest to order the columns in time chronological order according to the first year of each period.

- Why is there no entry for Zemp for 2003-2009. They have annual time series for all regions and the period means can easily be computed from their supplementary data.

- Why is Jacob et al. (GRACE) not included?

- WGMS 2015 is not in reference list? I am not sure this should be included here at all since (I assume) this is just a 'quick' extrapolation of the measurements rather than a thorough analysis as all others. The bias of the in situ measurements is known.

- For a reader it would be good to get some information how these estimates have been obtained since different methods have (sometimes known) biases. Perhaps a footnote reporting for each model in a few words what the source is (GRACE, extrapolation of observed in situ glacierwide specific balances, mean of in situ combination of XXX and XXX, mass balance model ...)

-