# Reply to the reviews of Paul Leclercq and Alex Gardner on our revised manuscript "Feedbacks and Mechanisms Affecting the Global Sensitivity of Glaciers to Climate Change"

We would like to thank Eric Larour for obtaining the reviews, and we would like to thank both reviewers for providing their detailed and helpful comments on our manuscript. We were able to address all their points (as detailed below).

## Response to Paul Leclercq:

*We reply to the reviewer's comments in italic letters below. But before we start, we would like to make a general comment on the nature of the reviewer's concerns:*

*All the points the reviewer makes concern exclusively the calibration and validation procedure of the model we use, which is described in great detail in Marzeion et al. (2012), such that we are under the impression of replying to a review of Marzeion et al. (2012), not the manuscript on hand. Marzeion et al. (2012) explicitly quantify, and show in dedicated figures, most of the error properties that the reviewer speculates on. Furthermore, none of the issues brought up by the reviewer has any potential to impact the results presented in the manuscript on hand, because (i) the uncertainties caused by them already are included in our uncertainty estimates, and (ii) Marzeion et al. (2012) explicitly show that these uncertainties are very small compared to the uncertainties caused by the forcing.*

*We also wish to point out something that the reviewer doesn't address at all: Marzeion et al. (2012) do not only analyze the uncertainty of the model quite carefully, but also do a completely independent validation of the error bars, such that in principle, we could give error bars on our error bars – i.e., we can be quite certain that we did not overlook any major source of uncertainty.*

*In the discussion below, we try to address the reviewer's concerns, but we think that it is not appropriate to include this discussion in the manuscript, because the uncertainties concerned do not impact the results, and they have already have been shown by Marzeion et al. (2012) to be minor in comparison to other uncertainties.*

*Finally, the reviewer in principle urges us to employ a more intuitive calibration procedure. We are not able to follow this advice, because we are not aware of an alternative procedure which would be superior, and indeed we have shown that some obvious alternatives are definitely inferior.*

### Calibration procedure

In the review I expressed my concerns about the calibration of the temperature sensitivity parameter mu*. The calibration is done by looking for the period t* with a climate in which the present glacier geometry is closest to balance. Demanding that for this climate the mass balance of the glacier with its present geometry is zero gives a value for mu*. For the 255 glaciers for which direct mass balance observations are available this equilibrium period is chosen based on the match between the modelled and measured mass balances. For the other glaciers, mu* is derived by spatial interpolation of the timing of this equilibrium period. I questioned the spatial interpolation of t* to derive the temperature sensitivity mu* for other glaciers.

In their response, the authors stress that they do not assume that the glaciers are in equilibrium in this equilibrium period t* , but that the glacier would have been in equilibrium in period t* if it would have had the present geometry. I already understood this (apologies if my comment was unclear in this respect), but I think it is good that this point is made more explicit in the revised version of the paper.

Furthermore, the authors indicate that by using their procedure they implicitly assume that both the the response time of glaciers and the climate fluctuations are regional characteristics. I think it is very reasonable to assume that climate fluctuations are regionally coherent, especially for temperature fluctuations which are most important for the mass balance of glaciers. Still it is questionable if a spatial interpolation of the timing of climate fluctuations makes sense: if a region A has a warm period in, say, 1920 and region B a warm period in 1970, is it than reasonable to expect that a region between A and B has a warm period in 1945? After all, the assumption might work because 1) for most regions quite a few measurements are available and 2) distance weighted interpolation is used, such that the calibration de facto never uses climate fluctuations of more distant regions. Nevertheless, the interpolation might be questionable in regions with very sparse mass balance observations, such as the Russian Arctic and Greenland.

*The reviewer's example of our method producing an undesirable interpolation between to "good" periods to a "bad" one is valid. In fact, it doesn't even need two different regions: Figure 1 illustrates that this danger even exists within small regions. Vernagtferner is situated in the Alps, where there is a high density ob mass balance observations. Figure 1 shows that there are two distinct periods that are "candidates" for $t^\star$ because they exhibit small biases when the glacier model is run under assumption of equilibrium with climatological forcing. Imagine that there is a second glacier with mass balance observations close by, which has a similar temporal development of the bias shown in Fig. 1, but the minimal bias (and therefore, $t^\star$) in the 1970s. For a third glacier situated between these two, and without mass balance observations, our method would then interpolate a value of $t^\star$ in the 1950s, while both glaciers with mass balance measurements actually indicate that this is a particularly bad choice of $t^\star$.*

*Therefore we agree with the reviewer that this is a problem – the big question is how to solve it best. While we don't have the answer, we are able to determine the success of different approaches to overcome the problem, because this is exactly what the cross validation is designed for: simply imagine three glaciers with observations in the example above, (which is easily the case in the Alps), and the cross validation will show bad results – i.e., large errors – for all three*

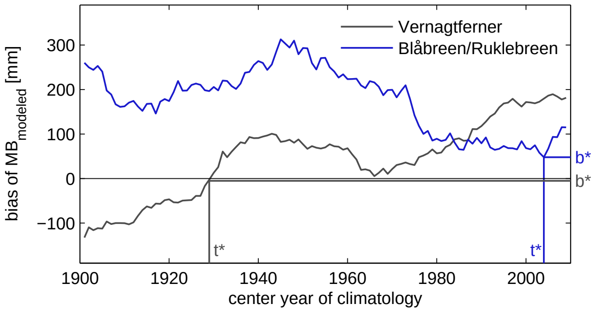


Figure 1: Two examples of potentially problematic cases for the identification of $t^\star$. Blåbreen is an example for a relatively large residual bias $b^\star$, Vernagtferner is an example where there are two distinct periods (around 1930 and around 1970) that could be identified as $t^\star$.

*glaciers.*

*When developing the model, we (i.e., Marzeion et al., 2012) tried many, many other methods then spatial interpolation – especially combining several "good" periods from neighboring glaciers, and trying to identify the maximum of their frequency distribution, etc. – but in the end, the simplest one (spatial interpolation with inverse distance weighting) turned out to produce the smallest errors.*

*So we can understand the reviewer's concern – we simply haven't found a better way to deal with it, and would welcome very much any contribution on how to improve it! That said: the errors resulting from the problem are included in our error estimate (see comment on cross validation above), and Marzeion et al. (2012) do indeed find a relation between the "remoteness" of the glacier and model error, as the reviewer speculates. There is a positive, significant correlation between the mean distance of the ten closest measured glaciers and both the root mean square error and the bias of our model – this is in Fig. 12c and d in Marzeion et al. (2012).*

*Finally: Marzeion et al. (2012) have explicitly shown that in spite of these uncertainties the spatial interpolation of* $t^\star$ *produces vastly smaller errors than the spatial interpolation of* $\mu^\star$.

Secondly, it is assumed that the response time of glaciers is a regional characteristic. The argument for this assumption is that the response time is dependent of the precipitation, which determines the mass turnover of glaciers. This is only partially true because also the glacier geometry is an important factor in the glacier response time. Up to a certain degree, the geometry is also a regional characteristic: the geometry of the Himalayas is very distinct from that of Svalbard. Nevertheless, differences between individual glaciers within a certain region can be large. For a previous paper (Leclercq and Oerlemans 2012, Global and hemispheric temperature reconstructions from glacier fluctuations) I have calculated response times for over 300 glaciers in different regions world-

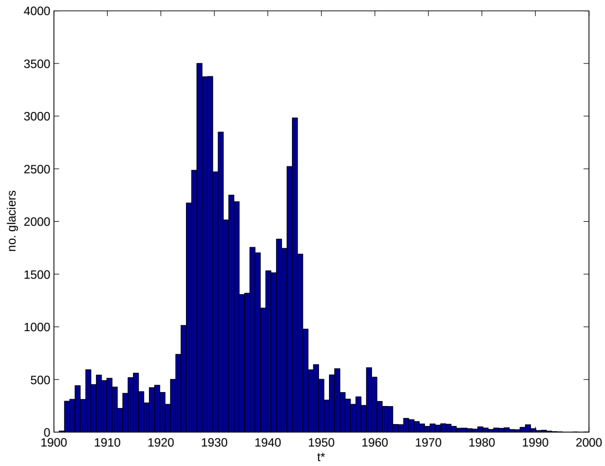


Figure 2: Histogram of the distribution of $t^{\star}$.

wide. Within regions differences in glacier response time of several decades (sometimes even more than a century) are found despite comparable amounts of precipitation. The authors state that the time between present and t* reflects the response time of glaciers (”timescale of response of the glaciers”). Given this differences in response time that do not necessarily follow a spatial pattern, it is quite a bold assumption that you can spatially interpolate t*, even if the assumption regarding climate fluctuations holds.

*We agree on all of this and can only repeat what we said before: Neither the model construction nor the calibration procedure is perfect, but we are able to (and do) quantify the uncertainties, and while it may be counter-intuitive that the spatial interpolation of $t^{\star}$ works, Marzeion et al. (2012) show explicitly that it is clearly superior to the more popular choice, which is a spatial interpolation of $\mu^{\star}$ (or equivalents).*

Furthermore, for many glaciers the response time is larger than 100 year, such that the timescale of response to the continuous climate change in the 20th century can be much longer than a century (the response time expresses the time needed to adapt to an instantaneous change in climate). This slow adaptation to past climate change is even part of the findings of this paper. As the CRU data set is only available since the beginning of the 20th century, the climate in which the present-day geometry is in equilibrium might very well not be included in the 109 available climatologies. How is this accounted for?

*As we say, $t^{\star}$ is not only a function of the response time, but also of the climate signal. Because the climate signal has been particularly strong during the past few decades, $t^{\star}$ is found well within the 20th century for most glaciers (see Fig. 2). But in principle, this is accounted for in the residual bias $b^{\star}$ (see Marzeion et al., 2012, and Blåbreen in Fig. 1).*

As said in the previous review, simplifications and assumptions cannot be avoided when modelling mass changes of all glaciers. However, by using the, what I called in the previous review, detour with t*, the authors use a rather complex procedure with implicit assumptions. I would prefer a more transparant, simple, calibration procedure. I am not confident that the assumption that climatic fluctuations and glacier response times can be spatially interpolated is reasonable. In their paper, the authors should make explicit mention of these assumptions that underlie their method. Furthermore, they should include the flaws in their assumptions in the uncertainty assessment.

*The uncertainties caused by the limitations pointed out by the reviewer are of course included in the uncertainty assessment, as argued above, and below. We also would prefer a calibration procedure (or in fact, a whole model construction) that is simple and gives better results - but we do not know of any.*

**Uncertainty estimate**

The second major point involved the uncertainty which is calculated using a leave-one-out cross validation. The resulting error for the global mass changes is surprisingly small given the uncertainties in the CRU data used for calibration of a glacier mass balance model (see e.g. Giesen and Oerlemans (2012) Calibration of a surface mass balance model for global-scale applications) and the simplified mass balance parameterization. I commented that the leave-one-out cross validation can only be used if the observations are truly independent. The authors gave an answer which I couldn't understand, but in an "off-line" discussion they explained it further. It turns out they answered a different, but not less important, question in the author reply.

My point was that the uncertainty in the mass balance of each of the 255 glaciers might have been underestimated. In their reply the authors focus on the calculation of the error in the global mass change, which is sum of the glacier mass balance. The errors in the mass change of the individual glaciers are independent because the global average of the correlation between the time series of the differences between modelled and measured mass balance (= model error) of the 255 glaciers with mass balance observations is practically zero. Therefore, the uncertainty in the sum is calculated as the square root of the sum of squared uncertainties in the individual mass balance records. The authors conclude that also at regional scale the errors are independent because the correlation of the error is not a function of the distance between glaciers (the "correlation of the correlation of errors with the distance between the glaciers is 0.008").

*We had the impression that the reviewer was mainly concerned about the size of the error bars of the global contribution, not that of the individual glaciers (i.e., if you look at Fig. 13 in the 2012 paper, the error bars probably don't look very small). Therefore we were assuming that the reviewer was questioning how we arrive at the very small global error bars from the relatively large error bars for each individual glacier. See below for a further explanation.*

This brings me to the following questions: Firstly, as most mass balance records are rather short, half of the records is shorter than 5 year, many of the 255 time series of the errors are also rather short. How long must a record be to make a significant conclusion about the independence of errors based on the correlation of the records? With two points the chance for a positive or negative correlation is more or less 50 - 50 %, right? And do you only take those parts of the records that overlap in time into account, or do you correlate the errors from different periods in time: e.g. correlate errors of one glacier in years 1978-1984 with those of 2002-2008 of another glacier?

*It is only possible to determine the correlation of the error during times when both glaciers have observations. This indeed leads to many relatively short error time series, but we do not test for significance nevertheless for two reasons: (i) There are very many pairs of glaciers ($\mathcal{O}(\frac{255^2}{2} \approx 30000)$) such that perfect positive and perfect negative correlations from 2-year overlaps will either average out (if random), or not (if there is a systematic correlation). (ii) If we would excluded non-significant correlations prior to calculating their mean, the resulting measure would not be able to tell us much about the error correlation we can expect for two arbitrarily selected modeled glaciers, but this is what we are interested in.*

Secondly, if you calculate the correlation between the correlation in error with distance between glaciers for all glaciers in the data set, then the correlation is dominated by the distant glaciers. Looking from one glacier, most other glaciers at large distances. For all those distant glaciers we expect little correlation between the errors which could then overrule the coherent behaviour of the few nearby glaciers. Is the correlation between error and distance also very small if you look at nearby glaciers only (distance $<$ 5000 km)?

*To see whether this is the case, we correlated the temporal correlations with the distance between the two glaciers. If the errors of close-by glaciers were correlated, but the errors of distant glaciers uncorrelated, you would assume to get a negative correlation (as is the case if you don't take the errors, but the mass balances themselves, i.e., close-by glaciers have more similar mass balances then distant glaciers). But we get a value of 0.008, indicating that the errors of close-by glaciers cannot be expected to correlate any better then the errors of distant glaciers. This is what is described on page 8, line 21ff of the manuscript.*

Thirdly, and probably most importantly, it is assumed that the small correlation between the errors of the 255 glaciers included in the calibration procedure implies that the error of all glaciers is independent. Are the calculated mass balances and errors of the other glaciers really independent of the results, and errors, of the glaciers that determine the model parameters?

*Due to the cross validation, we can be absolutely sure that the modeled mass balances that we use for the determination of the model uncertainties are independent of the observations. The same can not be strictly said for the model errors, since we assume that the mass balance measurements have no uncertainty – but this is a very theoretical limitation. We cannot be absolutely certain that the sample of $\sim 30000$ error correlations reflects the true error correlations between all glaciers – but we don't have any reason to believe otherwise.*

*Most importantly, we wish to point out something the reviewer hasn't addressed at all: Marzeion et al. (2012) do not only analyze the uncertainty of the model quite carefully, but also do a completely independent validation of our error bars, such that in principle, we could give error bars on our error bars. Also the results of this validation indicate that our analysis of model uncertainty is correct.*

## Response to Alex Gardner:

### Specific Comments

1. **Comment:** P4 L5-7: Do you have a reference to support this statement? At the regional scale, glacier mass changes in many regions have been directly linked to changes in temperature (i.e. Arendt et al., 2009 & 2013, Gardner et al., 2007 & 2011). Maybe I'm misreading this sentence.
**Response:** This really is an issue of the time scale considered – and at short time scales, there definitely is a strong link between glacier mass change and temperatures (after all, this is the basic assumption of our model). We reworded the sentence to clarify this, and added a reference to the Marzeion et al. (2012) and Leclercq et al. (2011) papers that give this result.

2. **Comment:** P4 L24-27: I found this sentence difficult to follow.
**Response:** There was an "of" missing - corrected.

3. **Comment:** P8 L3-4 Missing bracket
**Response:** The closing bracket is in line 6.

4. **Comment:** P8 L19-20: "Neighboring glaciers having very different temperature sensitivities". With all of my regional-scale analysis of glacier mass changes it seems more obvious that neighboring glaciers would have similar temperature sensitivities especially, outside of the mountain ranges of the Alps and High Mountain Asia. At the regional scale, glaciers have similar hypsometry and spatially correlated elevation changes. Can you point to any relevant literature to support your statement?

   **Response:** This can be understood as the effect of those cases where neighboring glaciers have different temperature sensitivities, although they experience a similar history of climate forcing. Also note that we are saying "having *potentially* very different temperature sensitivities" – i.e., the method is not at all excluding the possibility that neighboring glaciers have similar temperature sensitivities.

5. **Comment:** P8 L23-24: Remove second "correlation"
**Response:** Actually, this is correct - we are correlating the correlations with distance, so there need to be two "correlations" in this sentence (we are aware we are describing a complex measure – but we agree with reviewer one that we need to describe this somewhere, even though it would have been better in Marzeion et al., 2012).

6. **Comment:** P10 L26 Fractionation between solid and liquid -> fraction of solid to liquid precipitation
**Response:** Corrected.

7. **Comment:** P11 L5 and uncertainty of 0.2 cm SLE seems unrealistically small. You may want to revisit you calculation of the uncertainty.
**Response:** The small uncertainties are mostly a result of the large number of glaciers in combination with no correlation of the model errors,

neither in time nor in space (see also comment "P8 L23-24" above). This leads to the relative error decreasing with $n^{(1/2)}$, with $n$ the number of glaciers considered. Also note (as said in the next sentence) that these error bars are somewhat arbitrary nevertheless, since they depend on the length of the model integration, which for equilibrium experiments is a somewhat subjective choice.

8. **Comment:** P11 L6 Here you use 1 SE and on page 15 the authors use the 95 % confidence interval.
   **Response:** These are different numbers. We use the 95 % confidence on page 15 for comparability with the other studies' results.

9. **Comment:** P13 L10-12: Even after the author's explanation in his earlier response I find it difficult to see the utility of this experiment. Personally, I think the study would be much more concise without out it.
   **Response:** We agree that it is not realistic in any sense – but we think that the importance of this sensitivity case is that it is effectively what is assumed by models that use a constant mass balance sensitivity (kg $m^{-2}$ $yr^{-1}$ $K^{-1}$ averaged over the glacier) with V-A scaling (cited in the paper).

10. **Comment:** P16 L22 delete first "mass"
    **Response:** Corrected.

11. **Comment:** P17 L2-18 This paragraph would benefited from a rewrite to improve readability.
    **Response:** This paragraph combined two relatively unrelated discussions - we separated them into two paragraphs.

12. **Comment:** P17 L23 "But" -> "That said, "
    **Response:** Corrected.

13. **Comment:** P18 L4-13: My earlier comment that: "I agree that hypsometric feedbacks have contributed to 20th century rates of mass loss but I would be very cautious about making any broad conclusions as to why earlier rates of glacier mass loss were as negative later rates. I suspect that earlier estimates may be revised downward as speculated by Gardner et al. (2013)." I would be interested to here the authors thoughts on this.
    **Response:** Very sorry that we missed this in the first review!
    There are several reasons why we agree in principal that earlier estimates may be too high – one particular is related to the fact that a great fraction of the early 20th century glacier mass loss comes from a few years of exceptionally strong mass loss from peripheral Greenland glaciers. While there is anecdotal evidence that glaciers retreated strongly during that time, the temperature and precipitation time series we rely on are based on very few sparsely located observations. But there are two reasons why we think our conclusion as it stands is justified nevertheless:
    (i) It is based not on the glacier model forced by climate observations, but on the sensitivity runs forced by the "historical" CMIP5 runs - which do not generally show warm events around Greenland in a similar magnitude.
    (ii) There is evidence that the estimate of glacier mass loss from glacier length change observations (i.e., Leclercq et al., 2011) converges with the reconstruction in Marzeion et al. (2012) also in the early 20th century

when (a) additional length data, particularly from the Russian and Canadian Arctic and Greenland are included, and (b) the length change method is calibrated based on the most recent version of Cogley's (2009) mass balance observations, taking into account the RGI (Paul Leclercq, pers. comm.).

14. **Comment:** P18 L16 " . Particularly" -> ", particularly"
**Response:** Corrected.

15. **Comment:** P19 L18 "of the global ice mass in glaciers" -¿ "of glacier"
**Response:** Corrected.

16. **Comment:** P20 L7 "he" -> "the"
**Response:** Corrected.