Review of "The effect of overshooting 1.5◦C global warming on the mass loss of the Greenland Ice Sheet" by Martin Rückamp et al.

Summary

I am evaluating this paper for the second time after major revisions in the first round. The manuscript has changed considerably since the last iteration, has improved in response to the reviewer's comments and has seen new material being added. However, some major issues remain or have been introduced with the modifications, as is the case for the validation part and the "scenarios without overshoot". The language also needs further improvements to clean up errors and make the text clearer.

I will first respond to some of the author comments (blue), with my initial comments indented twice. General and specific comments on the new version follow below.

Response to the discussion

> - Yes, indeed this is an important point and we followed the reviewers suggestion. With using the parameters of Krapp et al. (2017) the direct output of the SMB from SEMIC has a misfit of about ~2m/a and a correlation of ~r2=0.5 by comparing SMB_RACMO_1960-1990 an dSMB_SEMIC_1960-1990 (almost similar for all GCMs used).

This evaluation should be extended and backed with figures to appear also in the manuscript. See general comments.

> - However, recalling Equation 3 and 4 from the manuscript, we do not use the direct output of SEMIC, but apply anomalies computed using SEMIC. The benefit of our approach is, that only the GCM trends of SMB changes are added to the RACMO SMB reference field, which represents the real SMB distribution very well. If we compare the computed SMB to RACMO (according to Eq. 3 and 4 without the synthetic SMBcorr), for instance for the HadGEM2-ES year 1990, it shows a very good agreement (Figure 2). See also answer to specific comment "p10 l2" below. In the revised manuscript we dedicate an own section to this issue.

This comparison does not really validate the SEMIC model. It only shows that at a given point (year 1990) the anomalies of SEMIC are close to zero. See general comments.

> - We expand the section about the SEMIC model in order to give the reader a better understanding of the model. In the new version of the manuscript we also review in

the introduction section briefly the already existing alternatives used and relate the discussion section accordingly. The reason we have not included too much detail on that issue previously is, that we basically apply SEMIC and that the model in itself and all the parameter tuning is work done by Krapp et al., 2017. The advantage of using a semi-complexity model is indeed its simplicity and cost efficiency, which would allow ice sheet modellers to also run computation up to time scales of thousands of years (e.g. until 5000) studying long-term commitment of various emission scenarios and hence not be limited by the availability of regional climate model output.

Multi-millennial simulations are not relevant for the current study. The question is if the chosen model is an appropriate tool for the presented type of simulation. The SEMIC model has so far not been validated for the use with GCM input data, the way you are using it for the projections. Showing a proper validation is the price you have to pay for that novelty.

- The authors rely on the parameter settings of the SEMIC model, which have been optimised for a different climate model input (Krapp et al., 2017). The Krapp et al. study shows that the SEMIC model can well approximate the MAR SMB results given MAR climate input. It must however be expected that the parameters that were chosen for a completely different climate input (different model, RCM vs GCM) are not optimal. Unless evidence can be provided that the applied parameters are indeed suited for the GCM forcing used in the present study, the model parameters should be optimised. Discussion on differences to other results (e.g. as done compared to Fürst et al., 2015) hinges on the implied sensitivity of the SMB model, which is currently not possible to be judged.

We haven chosen the same parameters of SEMIC as Krapp et al., 2017, due to the following reason: the parameter tuning procedure performed by Krapp et al., 2017 aimed to find a parameter set which gives a best fit between SMB and skin temperature Ts of SEMIC with only a limited number of processes and simpler parameterisations than a regional climate model with full complexity would derive. As a regional climate model is typically validated against reanalysis data and observations, the best match between SMB and Ts of SEMIC and regional climate model (in that case MAR) is the best way to represent the processes and their parameters in SEMIC. We see it thus as a tuning of the parameterisation of the processes. Once the process description in SEMIC is optimised, any type of input, either GCM or reanalysis data fields, will lead to the best possible SMB and Ts fields that SEMIC can produce. Still, the GCM will lack the best atmospheric fields over the ice sheet, as it is limited in resolution compared to a regional climate model. Given experiences we made from these three GCMs used in this study, which are all have different drawbacks, which would mean to have a tuning for each of them and this tuning would then make the whole benefit of having a semi-complexity model with low costs meaningless. Furthermore, it would basically mean to compensate far too low near surface temperatures with SEMIC parameters, which would offset the whole comparison of GCM forcing. Therefore, we have chosen a different approach: we compensate for this by using the SEMIC output only as an anomaly.

I understand the argument to avoid tuning the model for individual GCMs and I agree with the point on compensating errors in the GCM. Nevertheless, I think you will agree that if you were to tune SEMIC for another RCM (say RACMO or HIRHAM), you would end up with different parameters. I believe it is important to recognise that, even if you chose to do nothing about it and use the MAR based parameters.

The situation here is worse, because using forcing from a GCM implies different characteristics, like smoother gradients and less resolved geometry compared to the RCM. It is possible that these characteristic differences between RCM and GCM (not individual model bias) have an important impact on the modelled SMB. I believe it is your responsibility to show that the parameters that you are using are indeed appropriate for the given purpose. You should show how the absolute SMB looks like for the different GCMs and compare that to reconstructions and/or state-of-the-art RCM results.

I agree that using the anomaly method is a good choice, as it circumvents ***some*** of the biases in the absolute SMB products you are producing with SEMIC. Nevertheless, the reference SMB has an important impact on the results, because of feedbacks and non-linearities. I insist that you show the total SEMIC SMB somewhere (possibly in an appendix or supplement) so that the quality of the model can be judged.

- p6 l18 Not clear what the shortcomings of the Krapp method to treat albedo were and neither how this has been improved for the present study. This requires some additional description. Extending on the last comment, changes to the albedo scheme likely also have an impact on the SMB and would lead to different tuning even for the same climate model input.

We agree with the reviewer. We expand the section about the SEMIC model. In order to be consistent with parameters provided by Krapp et al. (2017) we switched back to the albedo scheme used by Krapp et al. (2017) for the new simulations.

So the improvement in the albedo scheme was not a very important improvement? As pointed out before, consistency may already be violated just by using a different climate model. Therefore, the consistency argument does not hold very strong for me.

- p9 l25 These gradients were found as best fit to SMB simulated by a specific RCM (MAR) at different elevations. Applying these in your setup may be better than nothing, but for a consistent picture, these should ideally be recalculated based on your own model setup(SEMIC). Maybe, if you can run SEMIC at different elevation, you could get a feeling for the implied differences. ***At the very least this inconsistency should be recognised and discussed as a shortcoming***.

This would be an interesting study. But for our application we follow the same argumentation above to the major point “parameter tuning”. The parameters found by Edwards et al. (2014) are the most physical reliable and additionally we don't want to have different parameters between the three GCMs.

I agree with the argument that having different parameters for the different GCMs is not desirable and I see that it would be extra work to recalculate them with SEMIC. I completely disagree with the notion that the gradients are "the most physical reliable". These calculations have since been made with other models (e.g. Noël et al., 2016) with clearly different results, which shows that these parameters are model dependent and not unique solutions. I iterate my minimum requirement to mention in the text that the gradients are based on a different model setup and not consistent with the climate forcing applied for the projections.

- p11 l32 I am wondering in how far a detailed analysis of individual glaciers is justified given that an important aspect of the forcing in form of interaction with the ocean and sub-glacial hydrology is missing. The comparison suggests that we could hope to get the behaviour of individual glaciers in line with observations, which I consider very unlikely given the steady-state initialisation, coarse GCM-based forcing and lack of important forcing mechanisms.

This is indeed a good point raised. It is certainly true, that important forcing mechanisms like the oceanic forcing and subglacial hydrology are missing in this study, however, representing the dynamics of a glacier in the narrow fjords of Greenland well or representing the large NEGIS well, is only achieved with sufficient grid resolution and physics in the model, which our model both fulfils. This is indeed assessed by comparing individual glacier drainage basins with observation, like the surface velocity field. We are concerned about the statement ‘given the steady-state initialisation’ – we do not perform a steady-state initialisation at all, in contrast, we perform a complex initialisation procedure with mixture between inversion and

paleo-spin ups. This procedure has been the top procedure in an international benchmark assessing the ability of models to achieve a good initial state (Goelzer et al., 2018). The reviewer seems to have overlooked this substantial part of this study. The coarse GCM-based forcing is subsequently processed in SEMIC is improving the resolution and the anomaly forcing is making sure, that the SMB in individual glacier basins is in high resolution – so the glacier basins are forced on high resolution.

With steady-state initialisation I mean that the attempt is to bring the ice sheet to a steady state at 1960. The way this is done here, no transient dynamical processes are active at that point that arise from past climate forcing. In the absence of dedicated ocean forcing, the response in ice flow and outlet glaciers can only be based on SMB forcing from 1960 onwards (and possibly some unwanted model drift). I believe this is also what the second reviewer had in mind for his second general comment on the omitting of ice dynamics.

The claim to have the "top" procedure in the initMIP benchmark calls for some clarification. The model clearly achieves a very good match with the observed geometry. However, this is not the only factor that should be evaluated to judge the quality of an initialisation. It is specifically pointed out in the Goelzer et al. study that for this class of models, a better match with the observed geometry can be achieved by accepting a larger drift in the control experiment. The model drift in the control experiment of ISSM-AWI is the ***largest*** in the group of models with a similar initialisation method (data assimilation). Taking isolated results of an intercomparison out of context to falsely claim a superior modelling approach is inappropriate and should be avoided.

General comments

The validation presented in section 2.3 has important problems:

The correlation analysis shown in Figure 4 is not a meaningful validation. The year-to-year variability in the GCMs is not expected to coincided with that of RACMO, because the GCMs have their own internal variability. Correlation other than the long-term trend is pure coincidence, as can be seen from figure 3.

The comparison presented in Figure 5 is also meaningless, because the two SMB fields are by construction very similar. Panel a is RACMO(1990) and panel b is mean RACMO(1960-1990) + dSMB, where dSMB is by definition close to zero. The difference between the two is only due to inter-annual variability in RACMO and the GMCs, which again, are not expected to co-evolve.

In a first step, the absolute SMB of SEMIC for the different GCMs should be shown and compared to state-of-the-art RCM results.

Secondly, a meaningful validation of the SMB model in response to climate forcing is to force SEMIC with reanalysis data (e.g. ERA interim) and compare the resulting SMB with observations, RACMO or MAR. This is done for any other SMB model used for projections, whether in absolute mode or anomaly mode (e.g. Hanna et al., 2011, Fettweis et al., 2007, Noël et al., 2016, Vernon et al., 2013).

Afterwards it can be concluded that the absolute SMB is not ideal and the anomaly method can be applied.

A new "scenario" (without overshoot) has been added to the analysis. It is constructed by "cut-and-paste" based on the original RCP2.6 simulations of the individual GCMs. The procedure first identifies the time of overshoot for each individual GCM. Until this point the forcing remains the same as for the original scenario. From this point on an arbitrary 30 yr period from later in the individual simulations is repeated to fill the length of the original simulation. I see a number of problems with this ad hoc approach that are mainly in relation to model dependency that complicates the comparison between the GCMs.

First, the resulting forcing time series should be shown. I suspect they will show a step change of the forcing at the moment of overshoot. I am not sure how to interpret such forcing as it is unphysical. It is also highly model dependent, I suppose the resulting forcing should not be called a scenario for that exact reason.

The choice of 30 year period seems arbitrary. Why not use instead e.g. the last 30 or 50 years of the simulation?. The strong multi-decadal variability visible in the SMB time series suggests that a much longer time period would be appropriate. How robust are results to a different choice of the period?

There may be a fundamental problem with the constructed time series because of using the anomaly method. The reference period for the SMB anomaly is 1960-1990, so the forcing is calculated relative to that period. The temperature time series used to diagnose the time of overshoot is referenced to another time period. This implies an offset of the forcing in function of the global temperature mismatch between the GCMs over the 1960-1990 reference period. I am not sure re-referencing will solve the problem entirely, but it may be worth a try.

I am not sure addressing these points will be sufficient to make the taken approach look like a good idea. To revert to the original manuscript and removing the constructed forcing may be a viable option, too.

Specific comments

p1 l11 Clarify the use of scenarios in "for some scenarios". Probably you mean "for some models" or "for some experiments" if is any of the 3 models and 2 scenarios.

p1 l11-12 Why "most likely"? How do the different experiments differ in terms of the integrated SMB? Is there a clear difference between the runs that stabilise and those that continue to lose mass after 2300? Is SMB integrated in time or spatially? Clarify.

p1 l14 Do you mean SEMIC or the GCMs in "stem from the underlying climate model"? Clarify.

p1 l17 Delete "observed" after "observed"

p2 l3-4 "mass loss" is caused by "acceleration" and **decrease** in SMB.

p2 l6 Maybe omit "regional" since the study is not concerned with it.

p2 l16 Also here, clarify the use of "other scenarios". See comment p1 l11.

p2 l19 place "has exceeded 1.5C ..." before "and may exceed 4C by 2100".

p2 l25 remove two times "very" before "scarce" and before "extensive".

p3 l6 Reformulate. "most suitable" may be true in some cases, but is certainly not generally true.

p3 l19 Replace "volume" by "thickness".

p3 l19 Remove "surface".

p3 l29 Add after or replace "Numerical" by "thermo-mechanical".

p4 l12-13 add "M" after "melt rate" add "R" after "Refreezing" and adjust text below.

p5 l4 It is confusing that the analysis is done with 11 yr moving windows and the lines in the figures are plotted with 30 yr running mean. This makes it difficult to visually inspect the threshold criteria and the location of the dots seems off. Consider revising.

p5 l14 Reformulate "striking". A large scale average will always show less variability compared to a local region in a dynamic system.

p5 l25 Conservative interpolation may not be optimal for temperature, a quantity that physically cannot be conserved. I suspect that the imprint of the original GCM grid in the final product we see in Figure 11 and 12 may be related to that. For a vertically downscaled variable, I would not expect such a strong imprint.

p5 l26 Insert "on which SEMIC is run" after "0.05 grid", if that is correct.

p5 l29 Remove line break, still discussing downscaling.

p6 l7 Add "when MAR is used as forcing" after "best possible SMB and T_s fields".

p6 l13 Add a figure that shows the absolute SEMIC output, e.g. the 1960-1990 average given in rhs of equation 5.

p6 l20 The integral of deltaSMB in equation 5 from 1960 to 1990 should be zero and the integral of SMB_clim over the same period should be the same for all GCMs. It doesn't look like that in Fig 3, but maybe that is because of the running mean? To check!

p7 l17 Move "integrated" after SMB to avoid confusion between spatial and temporal integration.

P7 l2 How far are you from the ideal case? Please show that as a figure plotting the difference between the two reference SMBs.

p7 l26-29 This text is not part of the validation. Suggest to move to the results section.

p7 l30 This is not a meaningful validation. See general comments.

p8 l10 This is also not a meaningful validation. See general comments.

p9 l1 The causality of this sentence is not clear to me. Why could an arbitrary time period not be used if absolute SMB would be applied? I believe there may be a fundamental problem arising from the use of the anomaly method. See general comments.

p9 l3 What is the motivation for using the time period (2250-2280), or is that an arbitrary choice? Please clarify in the text.

p9 27 How do you apply observed velocities to land-terminating glaciers? Only at the ice front, as a boundary condition? Please clarify.

p9 l33 This compensation only applies at marine margins, I suppose. Clarify.

p10 l1 Maybe "have retreated".

p10 l20 Mention which version of BedMachine.

p10 l23 Clarify what noise is expected to be avoided.

p10 l25 I think it is safe to replace "125 kyr before 1990" by 125 kyr BP, with “(before present)” in first occurrence, to avoid the confusion between 1990 and 1960 in the following.

p10 l30 Consider discussing the temperature spinup with constant climate after relaxation at p10 l23 and adding it to Table 3.

p11 9-12 This could be mentioned before describing the method. In any case, reformulate. I don't think you really assume these statements to be true: Replace 'assumptions' by "simplifications". "The currently observed present-day elevation is taken constant for the entire glacial cycle". "the basal friction coefficients obtained from the inversion is taken constant for the past glacial cycle, and (3) the temperature changes from the GRIP record are applied to the whole ice sheet without spatial variations."

p11 l20 I agree that it may be a negligible effect. But where is the table with comparison of basal temperature against ice core results? The suggestion was not to remove the table, but to check the results that were presented in it.

p12 l27 The drift is ~15% of the magnitude of the lowest projection. That's not negligible. Replace "negligible" by "small".

p13 l7-9 Figure 8a and 8b are confused. Exchange.

p13 l9 Explain blue dots in Figure 8a.

P13 l9 The number in the table rounds to 400 m/a not 390 m/a.

p13 l11 Was the model run forward in time here or in the comparison, clarify.

p13 l13 What is the "assumed critical time"? Clarify.

p13 l17 Replace "negligible" by "small" and give a number.

p13 l20 Replace "variability" by "range". This is an ensemble range.

p13 l21 Name which figure (Fig 9) after "mass change".

p13 l22-23 Better discuss only SLE change, otherwise it gets confusing with the different sign of mass and SLE changes.

p13 l29 The comparison with Fig 1 is hampered by the different reference periods used for temperature and SMB. Consider producing a temperature time series re-referenced to the 1960-1990 reference period.

p14 l3-8 You could mention already here why the numbers are expected to be lower (missing ocean forcing, missing Greenland blocking in GCMs, ...). It seems more appropriate to compare against SMB-only results for this period instead.

p14 l9 Shouldn't some of this section 3.3 appear before the projections in chapter 3.2, since it shows the forcing that leads to the ice sheet results?

p14 l16 Replace "cooling" by "less warming", or are you comparing 2300 to 2100? Not always clear what we compare against.

p14 l20-22 Clarify this contradiction: "amplification is not well represented in MIROC5" <?> “with respect to the Arctic amplification phenomena the most plausible distribution of surface warming is produced by HadGEM2-ES and MIROC5”.

p14 l21- p15 l5 The discussion of realism of future warming patterns remains very speculative and arbitrary.

p15 l2-5 The Watterson analysis is probably global, which may not be that meaningful for this Greenland application. This should be mentioned. MIROC5 often scores best when used with MAR compared to other GCMs. This could be mentioned, too.

p15 l21 Remove "increased" before thickening, or was it thickening already?

p15 l23 Not sure comparison with observations really holds here. You mean marginal thinning and central thickening as large-scale features? Yes, but the thinning must reach much further inland here. Consider revising.

p15 l25 Is there evidence from other studies that the 79 glacier is vulnerable? SMB changes in HadCM and MIROC seem very small here and the pattern of retreat looks almost identical. What happens in the unforced control run in this region? Are these changes in the figure also calculated relative to the control run (i.e. double differences)? They should! In any case it may be interesting to inspect and show the control run in a figure.

p15 l33-34 Here the discussion of 2100 and 2300 changes is mixed with recent changes.

p16 Maybe "ice discharge anomaly"?

p17 What is "It"? Clarify.

Figure 1 The offset between GCMS in temperature in (b) does not correspond with the curves in Fig 3 because of the different reference periods. This makes comparison later in the manuscript difficult and may also have an impact on the timing of the overshoot.

Figure 3 "according to Fig 4.". Replace "thick line" by "solid lines"

Figure 4 Remove or replace by a more meaningful comparison.

Figure 5 Remove or replace by a more meaningful comparison.

Figure 6 Are values trimmed at +-25? If yes, say so in the caption and give the min/max values.

Figure 8 Explain what the blue dots stand for in a. Does an r^2 =1.00 mean perfect correlation? How is that possible?

Figure 9 Replace "Straight" by "Solid"

Figure 10 May want to add an estimate for the SMB-only contribution.

References:

Fettweis, X. (2007). Reconstruction of the 1979-2006 Greenland ice sheet surface mass balance using the regional climate model MAR. The Cryosphere, 1(1), 21–40. http://doi.org/10.5194/tc-1-21-2007

Hanna, E., Huybrechts, P., Cappelen, J., Steffen, K., Bales, R. C., Burgess, E., et al. (2011). Greenland Ice Sheet surface mass balance 1870 to 2010 based on Twentieth Century Reanalysis, and links with global climate forcing. Journal of Geophysical Research, 116(D24), n/a–n/a. http://doi.org/10.1029/2011JD016387

Noël, B., Van De Berg, W. J., Machguth, H., Lhermitte, S., Howat, I., Fettweis, X., & Van Den Broeke, M. R. (2016). A daily, 1 km resolution data set of downscaled Greenland ice sheet surface mass balance (1958–2015). The Cryosphere, 10(5), 2361–2377. http://doi.org/10.5194/tc-10-2361-2016

Vernon, C. L., Bamber, J. L., Box, J. E., Van Den Broeke, M. R., Fettweis, X., Hanna, E., & Huybrechts, P. (2013). Surface mass balance model intercomparison for the Greenland ice sheet. The Cryosphere, 7(2), 599–614. http://doi.org/10.5194/tc-7-599-2013