The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2017-129-RC2, 2017

The Cryosphere
Discussions

Interactive comment on "Design and results of the ice sheet model initialisation experiments initMIP-Greenland: an ISMIP6 intercomparison" *by* Heiko Goelzer et al.

Anonymous Referee #2

Received and published: 18 August 2017

Review of Goelzer et al. "Design and results of the ice sheet model initialisation experiments initMIP-Greenland: an ISMIP6 intercomparison":

This manuscript describes the efforts of the ISMIP6 project to summarize various methods for initializing models of the Greenland Ice Sheet (GrIS) in preparation for experiments or projections of change. The impact of initialization is important as evidenced by previous ice sheet model inter-comparison (MIP) efforts, whose results were clouded by the 'uncontrolled-for' use of various initialization techniques. One impact of this is that initial GrIS conditions for projections across these earlier efforts varied widely, from conditions that (1) were constrained by the data assimilation and inversion techniques

to look almost like reality to (2) initial conditions that resulted from long freely-evolving equilibration simulations. This spread had a potentially significant impact on subsequent projections of sea level change.

Because of this issue I place value in this manuscript, which steps back to try and clearly illuminate the types of initialization techniques and their impact on perturbation experiments. I think as a survey of these techniques, this manuscript will be interesting to ice sheet modelers - including those who contributed to this MIP. In the following, my comments (which ignore minor grammar/editorial issues for now) are intended to (hopefully) improve the MIP by deepening analyses and interpretation of inter-model differences in the context of difference model choices (the main point of MIPs). In the few cases where I single out individual model results, be aware that I'm not criticizing these models/results per se, but only trying to use them to motivate a more general analysis. Finally I welcome any counter-arguments, if the authors feel I'm incorrect in any comments I've made.

Comments:

P3L12:"Changes in ice sheet geometry generally cause changes in atmospheric conditions over the ice sheet, and hence changes in SMB." Suggest that the word "feedback" be used here.

P3L13: "the most important effect is the height-SMB feedback" -> "an important effect is the height-SMB feedback" (other feedbacks aren't necessarily well-enough explored to make the 'most important' conclusion, in my opinion)

P3L17: for non-experts, perhaps also note for context the climatological fraction of total ice flux that comes from surface melt versus marine loss (instead of the fraction of current mass loss *trends* that arises from these two terms)

P3L27: "This is a very short period. . .": Suggest adding some interpretation as to why these divergent timescales make things difficult.

P3L32: These two subsequent paragraphs (short-term vs. long-term projects) describe an issue that is similar to the difference between weather forecasting (an "initial value problem") and climate prediction (a "boundary value problem"). The difficult intermediate timeframe in the weather/climate context (seasonal->decadal prediction) is similar to the intermediate decadal->centennial timeframe in the ice sheet context. In both intermediate timeframes, it becomes unclear which technique (or blend of techniques) to use. I suggest describing this analogy to place the problem facing ice sheet modelers in the context of similar problems from other fields.

P3L32: suggest reordering this section to more clearly separate discussion of 'long-term' versus 'short-term' simulations and the initialization techniques used for each. As it stands the discussion mixes long and short time scale simulations a bit too much.

P4L15: "limitations in observations" -> "uncertainties in observations"?

P4L17: "transferred to" -> "masked by alterations to. . ." ..?

P4L24: Suggest a brief statement "Given the wide diversity of ice sheet initialization techniques. . ."

P4L24: "The goal of initMIP-Greenland is to document, compare and improve the techniques used by different groups to initialise their state-of-the-art whole-ice-sheet models to the present day. . .". This implies initialization to a transient ice sheet state (given anthropogenic forcing). For example, it appears that Goelzer et al. (2012) find a $\sim$0.7m GrIS SLR commitment if CO2 concentrations are capped at present day levels. Other studies I am aware of also find similar levels of SLR commitment to present-day climate. This arises from both historical warming relative to the preindustrial Holocene (which is the climate state most consistent with the current GrIS geometry) and also future committed warming, in very roughly equal parts. I think a discussion of this point is critical in order to put the stated goal, of initialization of ice sheets under present-day forcing/assimilation, in context. As one thought experiment: if the above two papers are correct in their commitment estimates, a perfect spin-up to equilibrium under per-

petual present-day forcing should probably produce a present-day GrIS that is $\sim$15% smaller than that observed. Another: a perfect DA procedure should result in an initial ice sheet with a transient trend equivalent to the recent historical average trend. Thus: would it be safer to remove "to the present day" from the statement to retain generality with respect to the chosen 'time of initialization'? Or state that initialization to present day is a choice driven largely by availability of observational data, which is required in the case of DA/DAv initialization techniques?

Table 3: Some rows are identical in their entries. For example, DMI-PISM1 and DMI-PISM2. The difference in these types of rows should be identified, so the reader can hopefully identify why they produce different results.

P9L5: I think these differences in approaches are very important points. It would be useful to be able to cross-list the initialized states (e.g. in Figure 2) against the amount of 'constraining' that the modeling groups used to obtain that state.

Figure comment: several colors in scatterplot figures (e.g. Figure 2) are very similar. For example, ARC-PISM and UAF-PISM3. Even on a good computer screen it is very difficult/impossible to tell the difference. Is there any way to label each dot (or replace each dot with a letter or number, perhaps also associated with a unique color)?

General presentation comment related to above 2x points: allowing users to interactively explore the MIP results would be incredibly useful, if The Cryosphere would allow links to supplementary interactive public online plots (perhaps hosted to/linked via the ISMIP6 website). As just one example: for each online plot, being able to isolate the the DA/DAv/SP contributor 'pools' and allowing an interactive data tip to identify model names upon hover-over, would be fantastic. One reasonably easy way of doing this is Plotly, I'm sure there are others. https://plot.ly/python/line-and-scatter/, "Line and Scatter Plots" example.

Suggestion: would the contributors consider providing a mask of grid points on the *terrestrial margins* of their ice sheets that are (unphysical, at steady state) accumulation or (physically consistent) ablation zones? In my experience, it is possible for the terrestrial margin points from RCMs to potentially display positive SMB (presumably due to RCM bias). If a significant number of contributors to this MIP are using SMB fields that include these types of unphysical terrestrial margin accumulation zones, an important discussion could be: how these are handled and how the presence of this unphysical margin behavior could affect future simulations.

P9L15: “Again, a good match with the observed ice extent is more important than the SMB model itself.”: this statement, as a value judgement, is unclear to me.

P13L14: “However, it is noticeable that DA models that have been initialised with one data set show lower errors when comparing with that specific set of observations.”: this seems a reflection of a well-known statistical tenet, that one can’t use training data as validation data. Perhaps this similarity should be more obviously stated for readers. Put another way: to what extent is comparing against observations useful, when in many cases (DA/DAv) these observations may been already used as part of the initialization procedure?

Figure 4 (and general question): if an ISM is allowed to grow outside of observed boundaries, but receives SMB from an RCM that is limited in extent to the present-day ice ice extent, how is the SMB calculated outside of the present-day ice extent?

Figure 6: Models for which SMB is prescribed to be strongly negative outside the current ice extent will also show low RMSE thickness. I’m aware of at least one participating model that takes this approach. I would be surprised if other well-performing spin-up models did not do the same. Models that take such an approach should be identified so that their good RMSE thickness performance can be judged against the strong constraint of very negative SMBs outside of current ice extent (which will likely, by basic ice dynamics principles, translate into a pretty good ice sheet geometry).

General comment: it is difficult to assess the an unbiased spread of results when a few models are represented by multiple contributions. I’m not sure how one would remedy

this, but it leaves the reader a little confused - especially when the difference between some contributions from the same model is not immediately clear (see previous comment).

P16L16: “. . .relatively small mass change for most of them over the course of the 100-year experiment. . .”: when contrasted against observed/projected sea level rise rates, I suspect many of these rates of change, even after 100 years, are very large (for example, relative to realistic future GrIS SLR rates). One way to better show ctrl rates of change would be to show the anomaly of each model, relative to the initial volume (which is already represented anyways in Figure 2).

P16L19: “. . .when nine obvious outliers are ignored”: it seems that the role of a MIP would be to explain why these outliers exist instead of discarding them. For example, did these models use DA/DAv/SP?

P16L21: “In some cases of the ensemble, the modelled background trend arises from transient forcing of SMB and temperature over the past, but more often it is due to inconsistencies introduced during initialisation.” Is it possible to identify which are which (potentially via online plotting technique mentioned above)?

Figure8b: at what point in the simulations is the ‘mass trend in control’ (y-axis) assessed?

P17L21: “This relation arises because the prescribed SMB anomaly has been optimized for the observed geometry, but has not been limited to the observed ice sheet extent.” As I understand, this means that the prescribed SMB anomaly gets larger the farther one goes from the present-day margin (Appendix A). Thus, it seems the strong relationship between initial size (and closely related area) and sea level response is mainly a direct response to a design feature of the MIP, and is not an emergent property of the ISMs themselves. Because of this I question the initial volume/SLR response relationship the authors highlight.

P18L5: I appreciate that a real SLR projection would probably use the 'perturbation minus control' approach to remove drift. However, for the sake of a MIP, I feel that showing the raw results would be very useful, so readers can truly appreciate what each model is actually doing. Conversely, perhaps the authors could make a plot/table showing the ratio of drift to perturbation magnitudes for each model, so readers could better assess how significant the drift actually is (instead of assessing this via visual plot comparisons). For example, this would allow users to assess their favorite rule of thumb for acceptable drift (for example, mine is: drift magnitude must be 1/10 of the expected signal magnitude).

General: I think the supplementary figures may be of most interest to the participants in this MIP. I really appreciate their inclusion.

Figure S3: for models like AWI and ISSM whose elevation is not defined outside of the present-day mask, what happens to ice that flows out of the present day mask? Is it simply set to zero..?

Figure S3: suggest outlining the ice sheet margin in red, or another obvious color, so bare land vs. ice sheet can be distinguished.

Figure S4: It is clear from this figure which models allow their ice sheets to evolve outside of the present-day volume, and which do not. As above, I'd suggest noting explicitly in a table which models do so and which don't.

Figure S5 (subsetted in Figure 10): A reader will immediately question why the ctrl thickness change for many DA* models (AWI-ISSM2,AWI-ISSM1,IGE-ELMER1, IGE-ELMER2, UCIJPL-ISSM, MIROC-ICIES1, ULB-FETISH2) have such major dH/dt artifacts, and why these artifacts won't artificially impact SLR projections (by control-minus-perturbation or absolute change methods). In the MIP spirit, I'd like the underlying reason for this pattern described, and in close conjunction, also why other DAv-based models do NOT exhibit this pattern and why. I think there may be important lessons here. I'd suggest a section of text discussing this.

Figure S5 (subsetted in Figure 10): Similar to the above comment: several SP models display large ctrl trends (e.g. PISM1-6). This is surprising since the point of a spin-up is generally to remove such trends. Conversely, other SP models have much less of a trend. Along the lines of the previous comment, I'd suggest a section of text discussing why differences in ctrl trends for SP models arise. There could be additional important lessons here.

Figures S6/S7: The negative ends of these color ranges indicates (I think) that the SMB forcing is dominant over the dynamic response in these experiments (which exclude potential ocean-driven dynamical forcing). First of all, isolating the dynamic response is a nice addition here. Secondly, this is slightly interesting with respect to the overall MIP, since it indicates that the applied SMB anomaly is a very important player here (especially with respect to initial ice sheet size); ice dynamics are essentially second order. Perhaps if the authors agree with me they could add a discussion to this effect (or provide a counterargument).

P20L20: I think this is a valuable summary paragraph. Suggest again making an analogy to weather forecasts versus climate projections (and/or initial value versus boundary value problems).

P20L33: "potential artefacts introduced during interpolation": can the authors possibly provide a quantitative estimate of this source of error?

P20L37: "The large ensemble spread in sea-level contribution in the asmb experiment is mostly due to the extra ice in the initial ice sheet geometry." Yes, but as above I'd argue that this relationship is mostly a consequence of the design of the imposed SMB anomaly field. As the authors note, a more realistic anomaly field that takes into account ice sheet geometry bias (e.g. a lapse-rated PDD scheme or an EBM scheme on multiple elevation classes) would likely show a much less pronounced volume-response relationship. So as stated, it's not clear this is a problem to focus community efforts on exclusively (though, of course the less initial volume bias the better, as long as the bias reduction technique doesn't deleteriously impact subsequent projections).

General: thanks to all the participants/coordinators of this MIP. This was no small effort.

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2017-129, 2017.