Interactive comment on "Glacial cycles simulation of the Antarctic Ice Sheet with PISM – Part 2: Parameter ensemble analysis" by Torsten Albrecht et al.

Response to Anonymous Referee #1 (Received and published: 26 June 2019)

We would like to thank the reviewer for the very constructive criticism regarding our manuscript. We were able to address all requests, find in **blue** the referee's comments in in **black** the author's response.

1 Overall assessment

This study presents a large ensemble modelling of the Antarctic ice sheet over the last two glacial cycles with the PISM ice-sheet model. The ensemble reveals clusters of best fit parameters that are evaluated against a series of constraints related to the present-day ice sheet and glacio-geological evidence. Results of the best fit(s) reveal the deglaciation history of the Antarctic ice sheet (in line with previous results reported in Kingslake et al., 2018) and show the major ice loss after MWP-1A.

My first concern is the choice of the sensitivity parameters in the ensemble, which is limited to four factors: ice softness (ESIA), sliding plasticity (PPQ), precipitation scaling (PREC) and mantle viscosity (VISC). Similar studies also explore the sensitivity of sub-shelf melting and how it relates to changes in far-field/continental shelf ocean and salinity in terms of oceanic forcing. Especially with respect to the explanation of MWP-1A, ocean forcing and its relation to sub-shelf melt may have played a crucial role (Golledge et al., 2014). Similar sensitivities have been explored in Pollard et al. (2016). Why are such parameters not taken into account, both sensitivity parameters within PICO, but also sensitivity to forcing, i.e. relation between atmospheric/ocean temperature forcing, for instance? As I understand from the paper, ocean temperature/salinity changes in the far field are not considered, neither through an offline ocean model, nor a parameterization that links atmospheric temperature change to oceanic temperature change. This is extremely important, as the conclusions with respect to the deglaciation do not take into account this sensitivity, hence show a large deglaciation pulse significantly later than the occurrence of MWP-1A. Many studies have shown the importance of the ocean in the dynamics of the Antarctic ice sheet, but neither the sensitivity (of PICO) or any ocean forcing has been investigated.

We thank the reviewer for pointing out important aspects of the parameter choice and implications for the last deglaciation. Some of these questions are actually touched in the first part of the study (Albrecht et al., 2019a), which certainly need to be better referenced in this second part. Find our detailed comments below.

A related question is why choosing those four parameters (ESIA, PPQ, PREC and VISC) and not others? Have other studies or previous experience shown that these are the most sensitive/critical? Some explanation should be given.

The choice of the four ensemble parameter is motivated in the companion study (Albrecht et al., 2019a), in which different parameter choices and boundary conditions are compared to a reference model ice volume history to gain some „prior model experience" and to determine most relevant parameters (for this metric) in each of the different model components (climatic forcing, basal sliding, ice creep and bedrock response). We agree, that this parameter choice is somewhat biased to the modeled total ice volume at LGM and present-day state, while other parameters may be more relevant for the onset and

pace of deglaciation. We have added a paragraph on deficiencies of the study. Yet, enhancement factors, sliding coefficient and viscous relaxation time of the bedrock have been also typically varied in previous model studies (e.g., Pollard and DeConto, 2012; Maris et al., 2014; Quiquet et al., 2018). As PISM uses a more generalized basal sliding and bedrock deformation scheme, we have selected different uncertain parameters.

Regarding the reviewer's concern on the sub-shelf melt sensitivity and the MWP-1A, we can state that Pollard et al. (2016) was focussing mainly on ice-oceanic deglacial processes in the WAIS with other relevant parameters fixed, while we consider a broader range of sea-level relevant processes over a longer time scale, such as ice-internal and ice-atmospheric effects, covering both parts of the Antarctic Ice Sheet. Golledge et al. (2014) used an apparently more realistic ocean forcing (from an Earth System Model), but they state „that there is considerable uncertainty in the relationship between ocean temperature and ice-shelf melt". In fact, much of the oceanic uncertainty of previous models is considerably reduced in our PISM simulations as it uses the PICO module (Reese et al., 2018), in which two uncertain parameters have been constrained by observed melt rates. Of course, we do consider ocean temperature changes, in our case coupled to surface temperature forcing (see Sect. 4.3 in Albrecht et al., 2019a). However, this relationship cannot account for events such as the Antarctic Cold Reversal after MWP-1A, when surface and intermediate water temperatures became rather decoupled. Yet, we have tested our PISM-PICO model for an earlier warming signal in the deeper ocean layers (while the surface was warming at the same time) in the companion paper (see Sect. 5.2 in Albrecht et al. 2019a), which can cause earlier retreat, while we still do not find main deglaciation before MWP-1A. For these reasons we have selected PREC as climate uncertainty instead of an ocean-melt (or calving) related parameter, as it can potentially counteract the other more constrained climatic forcings (see Sect. 4.5 in Albrecht et al. 2019a), and thsi aspect may have been underestimated by previous model studies. We have added some more discussion on the limited parameter choice and consequences for the results in the revised manuscript.

A second concern is about the novelty of the study, that methodologically is heavily relying on Pollard et al. (2016) and is basically performing the same analysis. However, a clear rationale on the choice of the boundaries for the parameter changes is lacking. Moreover, as shown in Figure 4, clear clusterings in misfit show up and best fit results are generally found in a much smaller range of parameter values (basically the range of two values for each parameter. Therefore, it seems to me that a smaller range sub-sampling would lead to an improved fit, hence reduce the uncertainties of the whole ensemble.

Yes, we have been using very similar analysis (and visualization) tools as discussed in Pollard et al. (2016) to allow for better comparison. However, Pollard et al. (2016) used different parameterizations in their model and focussed mainly on ice-oceanic processes in the WAIS over the last 20kyr. We have hence chosen different parameters and parameter boundaries as motivated in the (first part) companion paper (Albrecht et al., 2019a). A refined analysis could likely provide better constrained (best fit) parameter ranges. But the high uncertainty in the sea-level history is in fact a result of multiple best-score parameter ensemble members which show quite different sea-level histories.

Also, the best-fit parameters of our paleo study might be shifted slightly for higher spatial resolution, e.g. when performing short-term projections. In this analysis we wanted to consider the broader range of parameter values, covering the wide parameter range used also in other models (implying a rather coarse sampling) to gain a better understanding of

(combined) parameter effects in the highly nonlinear model. This also serves as rough constraint for further ensemble simulations and projections with PISM.

2 Specific remarks

Line 153: in -> to

Thanks.

Line 180 and following: All scores are aggregated into one score, thereby giving them an equal weight. However, some constraints are more reliable than others. Would different weighing lead to different results? Is there a certain bias towards one or several parameters; in other words, what is the result if scores would be calculated separately? Does this lead to the same clustering? Which scores are more representative?

This is definitely true, the score aggregation hides lots of information on the individual data types. However, we did not use inter-data-type weighting, e.g. based on spatial and temporal volumes of influence of each data type, as done in previous studies (Briggs and Tarasov, 2013; Briggs et al., 2014). Here, we followed the arguments in Pollard et al. (2016), assuming that „each data type is of equal importance to the overall score, and that if any one individual score is very bad ($S_i \approx 0$), the overall score S should also be ≈ 0... if any single data type is completely mismatched, the run should be rejected as unrealistic, regardless of the fit to the other data types... The fits to past data, even if more uncertain and sparser than modern, seem equally important to the goal of obtaining the best calibration for future applications with very large departures from modern conditions". We will refer more clearly to these argumentation in the revised manuscript.
We have also added an Appendix B with plots of individual paleo data misfits analogous to Briggs et al. (2014). If using the inder-data-type weighting and defining the score as weighted sum as in Briggs et al. (2014), the resultant distribution of best scores (here the smallest values) actually turns out to be very similar as for the product of individual scores, as shown in Fig. R1.
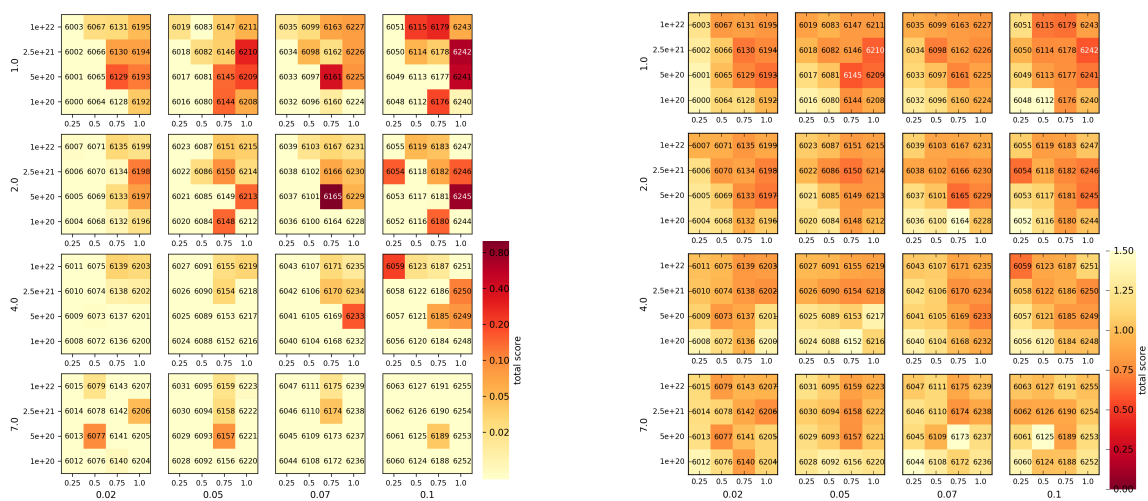

Fig. R1: Aggregated scores as product of individual scores as in Pollard et al. (2016) and used in our study (best fit equals 1, log color scale), compared to the aggregated score as a result of a inter-data-type weighted sum, as in Briggs et al. (2014), with best fits for lowest scores.

In fact, we can learn more about the model's response when discussing statistics on individual scores. Some of these information can be estimated from Fig. 2 or Fig. 5 and are discussed rather qualitatively in corresponding sections. We added ensemble standard deviation (Table R1) for each data type and some more discussion on the stastical aspects to the revised manuscript. Find also Fig. R2 and Fig. R3 for comparison (analogous to the ones in the manuscript, but separated for individual data-type scores).

	TOT	TOTE	TOTI	TOTDH	TOTVEL	TOTGL	TOTUPL	TROUGH	ELEV	EXT
MAD	0.002	0.123	0.023	0.183	0.144	0.099	**0.292**	0.156	0.049	0.047
SD	0.082	0.156	0.035	0.190	0.179	0.126	**0.300**	0.204	0.075	0.072

Table R1: Medan absolute deviation (MAD) and standard deviation (SD) for each data-type score, SD values are used in the revised manuscript.

In the manuscript (Sect. 3.1) we have discussed for each ensemble parameter how best scores are related to individual data types, as shown in Fig. R2. We want to avoid additional figures in the manuscript, but we added a general comment to Appendix B:

„The corresponding variability of each of the resultant normalized scores hence contribute different skills to the aggregated score (see Table 2). Generally, grounding-line related (TOTE, TOTGL, THROUGH) and ice volume-related data-types (TOTDH) show similar individual score patterns (not shown here) with ensemble standard deviations of 0.1-0.2. In the aggregated score this patterns becomes even more pronounced, while paleo scores (ELEV and EXT) and ice shelf extent (TOTI) show only little variation (<0.1) among the ensemble, and hence only little effect in the aggregate score pattern.“
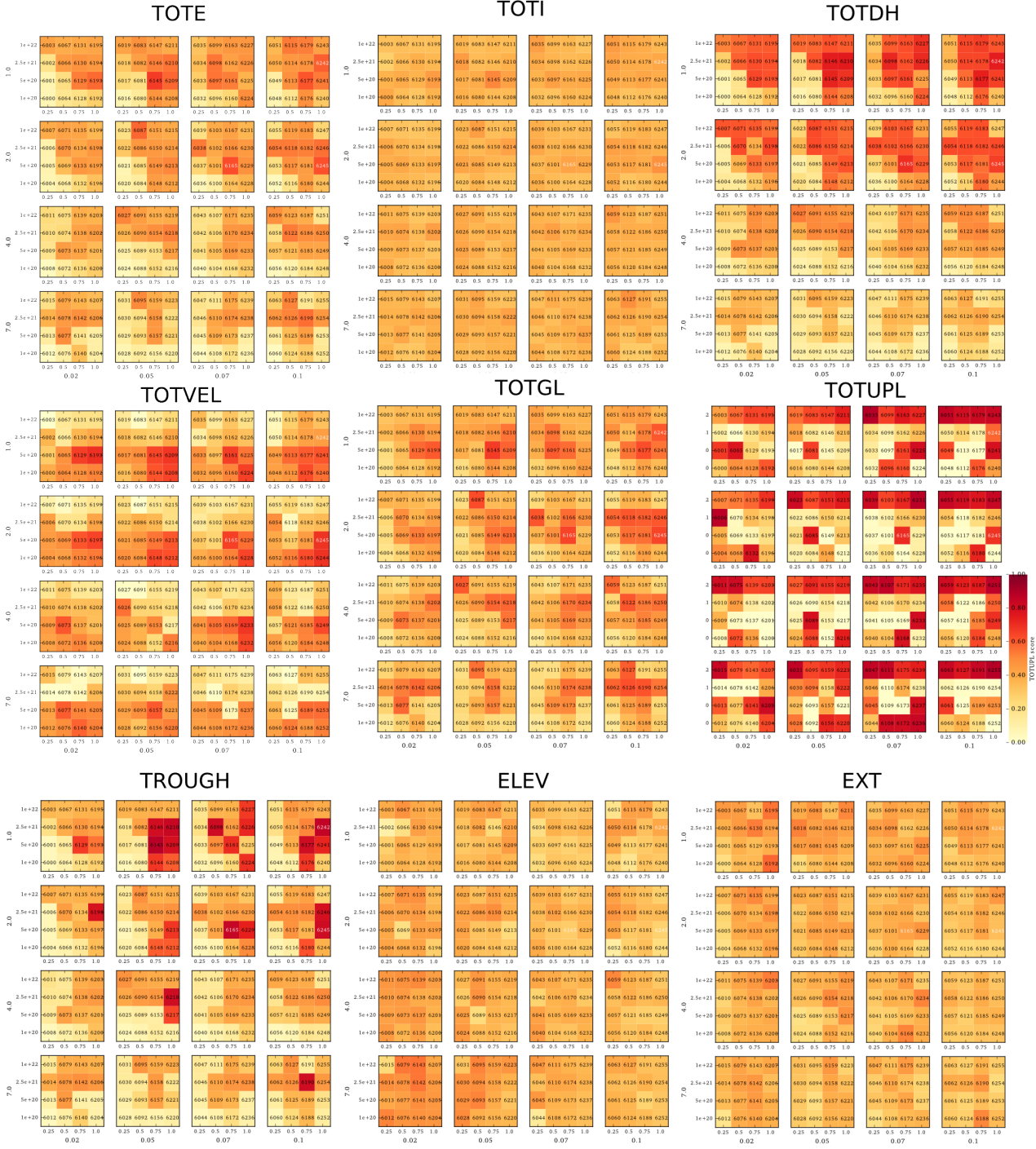
Fig. R2: Individual data-type scores for all 256 ensemble members, as in Fig. 1 in the manuscript, but with linear color scale. Scores in individual data-types are normalized by median.
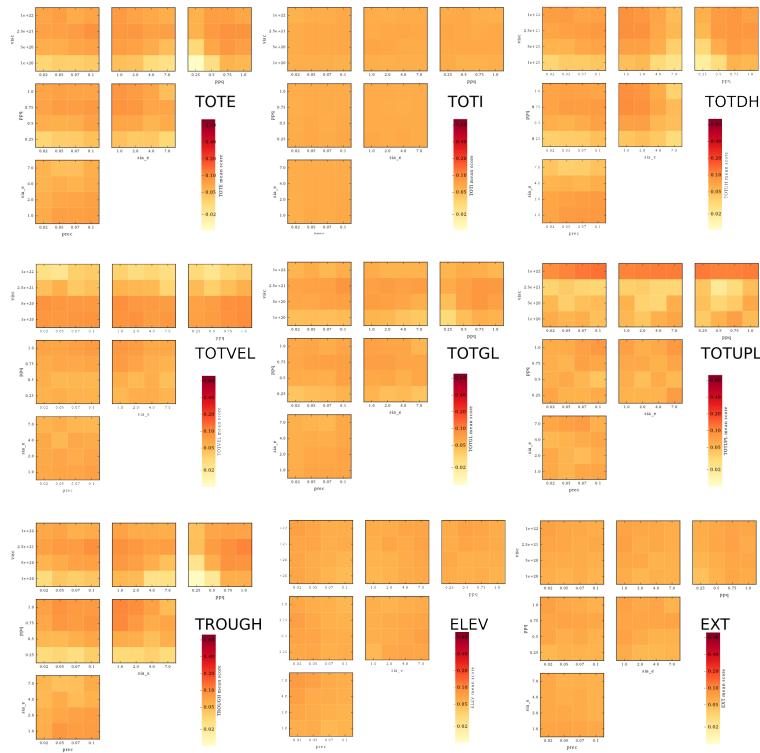
Fig. R3: Individual data-type mean scores for six possible pairs of parameter values, as in Fig. 4 in the manuscript.

The score pattern is also shaped by the uplift-related individual score (TOTUPL), that shows the highest ensemble standard deviation of 0.3 (Table R1) with a clear tendency towards higher VISC values, (see Fig. R3) probably a result of lower sensitivity to fluctuations in grounding line location. In contrast, the velocity-related individiual score (TOTVEL) with ensemble standard deviations of 0.2, favors lower VISC values, probably a result of more advanced grounding line location, which implies lower ice shelf velocities and hence lower chance for misfit (see Fig. R3). In the product formulation of the aggregated score such a reverse pattern can lead to highest total values for intermediate parameter values for VISC. We have discussed this aspect in the revised manuscript:

„As mantle viscosity determines the rate of response of the bed to changes in ice thickness a low viscosity corresponds to a rather quick uplift after grounding line retreat and hence to a retarded retreat, which corresponds to a rather extended present-day state. This implies smaller ice shelves with slower flow and less velocity misfit, such that also TOTVEL favors small VISC values. In contrast, a trend to rather high mantle viscosities in the aggregated score stems mainly from the misfit of present-day uplift rates expressed as data-type score TOTUPL, probably due to reduced sensitivity to fluctuations in grounding line location. High mantle viscosities involve a slow bed uplift and grounding line retreat can occur faster. More specifically, in the partially over-deepened ice shelf basins, which have been additionally depressed at the Last Glacial Maximum by a couple of hundred meters as compared to present, grounding line retreat can amplify itself in terms of a regional Marine Ice Sheet Instability (Mercer, 1978; Schoof, 2007; Bart et al., 2016). In fact, the best score ensemble members are found for intermediate mantle viscosities of VISC =0.5×10^{21} Pa s, and VISC =2.5×10^{21} Pa s. This could be a result of the product formulation of the aggregated score, in which individual data types scores
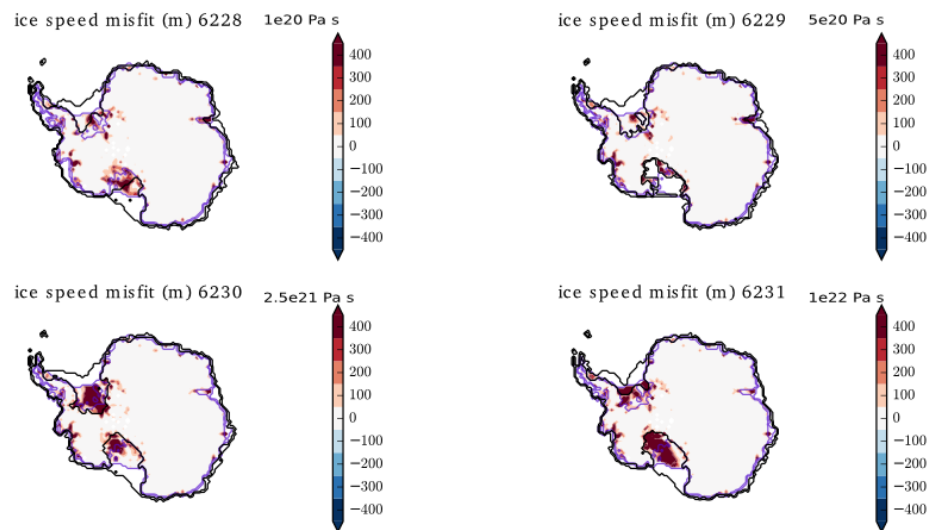
favor opposing extreme values."

Fig. R4: Map of misfit of modeled modern surface velocity (related to TOTVEL) in four ensemble members with different VISC values indicated in labels (but otherwise identical parameters).

Line 256: intermediate values of mantle viscosity give the best results. However, these are values for the whole Antarctic continent and several studies show that there is a distinct contrast in mantle viscosity between WAIS and EAIS. Would this not explain the best score (mean of both extremes)?

This is a good question. Recent literature suggest comparably small values for the oceanic WAIS plate. As most of the ice volume and grounding line changes occur in WAIS, one would suggest that this regions also leaves the strongest imprint on the individual-scores, that are related to the VISC parameter.

As already mentioned above, in our ensemble it is the TOTVEL data type which favors lower values, likely related to the grounding line location and ice shelf extent (see also Fig. R4), while TOTUPL actually favors large VISC values, which might actually be related to better scores for the EAIS part, where lower bedrock sensitivity and lower measurement uncertainty leads to lowest misfits (see Fig. R5).
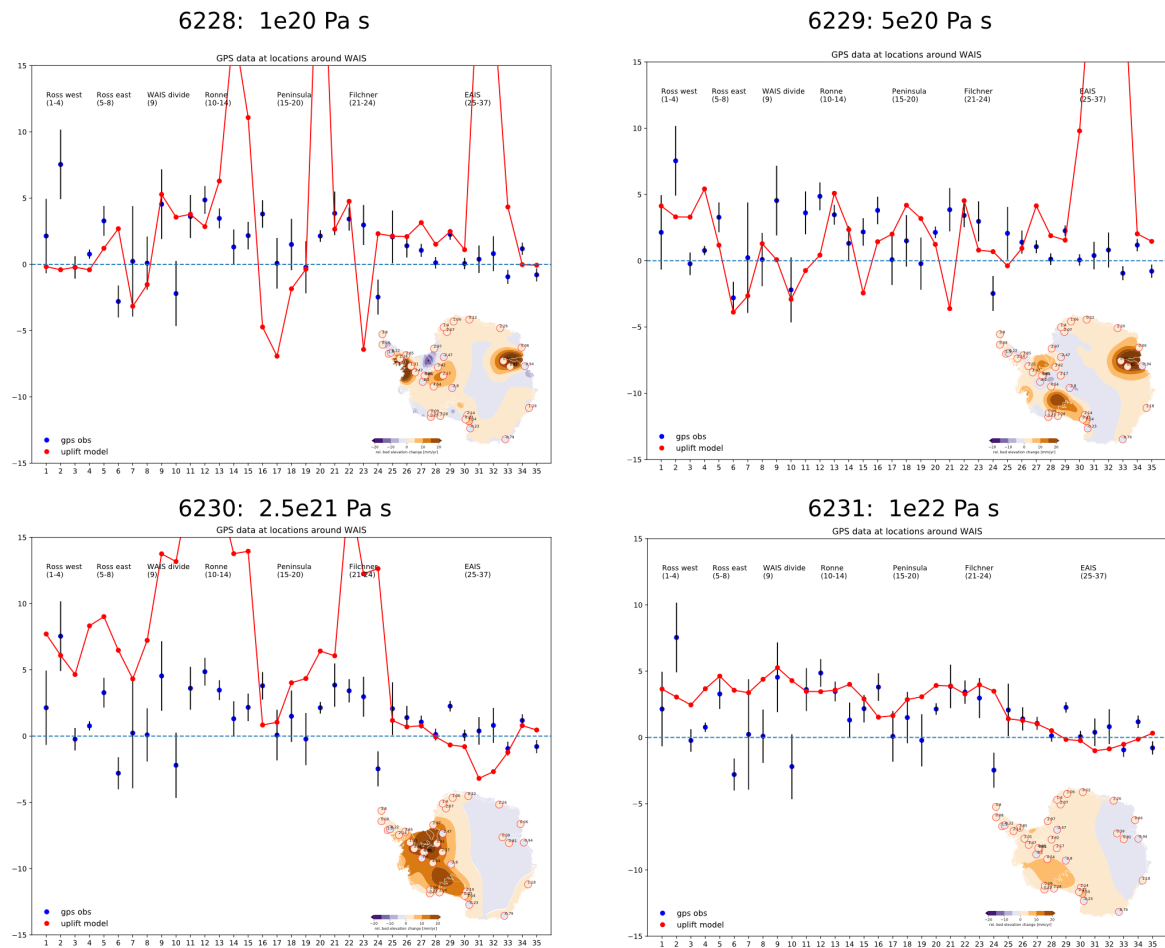
Fig. R5: Misfit of modeled present-day bedrock change rates to GPS measurements (related to TOTUPL) around the Antarctic continent for four different VISC values. Insets show location of PGS sites and map of bedrock change.

Line 278-79: why high basal friction? The power of the friction law only determines how sliding scales with τb .

Thanks for pointing out this imprecise formulation. Basal shear stress τ_b balances the driving stress within the SSA stress balance. As in the PISM friction law u_0 is considered as reference velocity (Eq. 2), such that for $q>0$ slower flowing upstream regions experience reduced basal shear stress, while fast flowing regions downstream are subjected to increased basal shear stress. Thus, reducing q from 0.75 to 0.25 produces slower flow in the interior and faster ice stream flow. We omitted this confusing aspect in the paragraph in the revised manuscript:

„In about 10% of the score-weighted simulations grounding line remains at the extended position without significant retreat, linked to ~~high basal friction (PPQ=0.25) and~~ an efficient negative feedback on grounding line motion related to a fast responding bed (low VISC)."

Figure 4: see general remarks: clustering demonstrates that the sampling range is too large and can be refined.

As already stated above we intended to cover a broad range of parameter values typically (and plausibly) used in other ice sheet modeling studies for better comparison (ESIA, PPQ, VISC) and to gain a better understanding of the actual (combined) effects of

parameters on the ice sheet dynamics. For follow-up projections (with higher resolution) a similar score scheme may be used, but for different (more recent) data-types in terms of hindcasting, with more refined parameter ranges.

Line 334: sub-surface melt: ambiguous, could point to melt occurring just below the surface. Using a term as sub-shelf melt is more appropriate.

Thanks, has been changed accordingly in the revised manuscript.

Line 378: remove 'with' and add year of communication.

Changed to „(personal communication Dave Pollard, 2017)".

References

Albrecht, T., Winkelmann, R., and Levermann, A.: Glacial cycles simulation of the Antarctic Ice Sheet with PISM – Part 1: Boundary conditions and climatic forcing, The Cryosphere Discuss., https://doi.org/10.5194/tc-2019-71, in review, 2019.

Briggs, R., Pollard, D., & Tarasov, L. (2013). A glacial systems model configured for large ensemble analysis of Antarctic deglaciation. *The Cryosphere*, 7(6), 1949-1970.

Briggs, R. D., & Tarasov, L. (2013). How to evaluate model-derived deglaciation chronologies: a case study using Antarctica. *Quaternary Science Reviews*, *63*, 109-127.

Briggs, R. D., Pollard, D., & Tarasov, L. (2014). A data-constrained large ensemble analysis of Antarctic evolution since the Eemian. *Quaternary Science Reviews*, *103*, 91-115.

Golledge, N. R., Menviel, L., Carter, L., Fogwill, C. J., England, M. H., Cortese, G., & Levy, R. H. (2014). Antarctic contribution to meltwater pulse 1A from reduced Southern Ocean overturning. *Nature Communications*, *5*, 5107.

Maris, M., De Boer, B., Ligtenberg, S., Crucifix, M., Van De Berg, W. J., & Oerlemans, J. (2014). Modelling the evolution of the Antarctic ice sheet since the last interglacial. *The Cryosphere*, *8*(4), 1347-1360.

Pollard, D., and R. M. DeConto. "Description of a hybrid ice sheet-shelf model, and application to Antarctica." *Geoscientific Model Development* 5.5 (2012): 1273-1295.

Pollard, D., Chang, W., Haran, M., Applegate, P., and DeConto, R.: Large ensemble modeling of the last deglacial retreat of the West Antarctic Ice Sheet: comparison of simple and advanced statistical techniques, Geosci. Model Dev., 9, 1697–1723, doi:10.5194/gmd-9-1697-2016, 2016.

Quiquet, A., Dumas, C., Ritz, C., Peyaud, V., & Roche, D. M. (2018). The GRISLI ice sheet model (version 2.0): calibration and validation for multi-millennial changes of the Antarctic ice sheet. *Geoscientific Model Development*, *11*(12), 5003.

Reese, R., Albrecht, T., Mengel, M., Asay-Davis, X., & Winkelmann, R. (2018). Antarctic sub-shelf melt rates via PICO. *The Cryosphere*, *12*(6), 1969-1985.