Glacial cycles simulation of the Antarctic Ice Sheet
with PISM - Part 2: Parameter ensemble analysis
Torsten Albrecht 1, Ricarda Winkelmann 1,2, and Anders Levermann 1,2,3

This part II submission of Albrecht et al examines a moderately sized ensemble of Antarctic glacial cycle runs with the PIK variant of the PISM ice sheet model. The ensemble runs are scored against a range of paleo and present-day (PD) constraints. The scored ensemble uses a reasonably state of the art model for ensemble glacial cycle contexts and adds to the litterature of what various models will do for past AIS glacial cycles. I therefore see the study potentially worthy of publication in TC given the current bar. At some point in the future I hope that model-based studies will have the requisite level of uncertainty quantification to enable much more meaningful inferences about past ice sheet evolution. However even with the current bar, a number of significant deficiencies need to be addressed.

The experimental design has some significant problems that are not even discussed. The study only using 4 ensemble parameters. Briggs et al, (TC, 2013) for instance, have 12 ensemble parameters just for the climate forcing and 31 ensemble parameter in total. At least 1 of the 5 temperature related ensemble parameters in that study (Tmix1) was one of the most sensitive ensemble parameters (with generally more sensitivity to this than to the precipitation related parameter (PdeselevEXP) that best corresponds to the sole climate forcing parameter (PREC) in this submission. The lack of an ensemble parameter relevant to the temperature forcing is especially problematic given the stated context of providing a distribution of present-day ice sheet states for initializing future projection runs. That state will depend significantly in the 3D temperature field of the ice sheet, the uncertainty of which is not probed in this study. Ideally this would be remedied, but that would be a major endeavor. At the very least I expect a clear and complete discussion of model and experimental design weaknesses and associated relevance to given results. A summary of this should also be in the conclusions.

Another major omission is a comparison of the ensemble results against the paleo data constraints. The chosen normalization of all score components against median scores means that the scores do not convey any information about absolute model fit to paleo data, only relative fit. It is therefore incumbent that the complete set of ensemble fits to paleo constraints be explicitly shown and discussed, eg as done in Briggs et al, 2014.

Furthermore, there are a number of claims and statements (detailed below) made that I find are indefensible, misrepresentative, and/or incorrect.

Specific comments

Large Ensemble of 256
-> Ensemble of 256
with eg Briggs et al using a 2000 member ensemble, you can hardly call 256 a
"Large Ensemble".

"The model is calibrated against..."

-> The model is scored against...
scoring a moderately sized ensemble is not calibration

what is the SSA enhancement factor value? It is never explicitly given
in this study nor in the PART I of this submission.

This analysis further constrains relevant model and boundary parameters by revea
ling clusters
of best fit parameter combinations.
Isn’t that already previously stated in different words:
"The model is calibrated[scored] against..."

Our Large Ensemble analysis also provides well-defined
parametric uncertainty bounds and a probabilistic range of present-day states th
at can be used for
20 PISM projections of future sea-level contributions from the Antarctic Ice She
et.
Kind of meaningless. I can think up a dozen metrics that would provide "well-
defined
parametric uncertainty bounds", each with different resultant ranges.

nonconserving
sub-glacial hydrology model
as from review of part I: pretty crude to call this a model ->
Here we use the non-conserving sub-glacial hydrology parametrization

Missing brief (eg 1 sentence) description of bed thermal model.

Sub-shelf melting in PISM
is calculated via PICO (Reese et al., 2018) from salinity and temperature in the
 lower ocean layers
on the continental shelf (Schmidtko et al., 2014) in 18 separate basins based on
 (Zwally et al.,
2015) adjacent to the ice shelves around the Antarctic continent
the companion paper states that salinity was not varied:
"While salinity change over time in the deeper layers is neglected in this stu
dy"
and this should be made clear here.

use the Large
Ensemble approach
Why is this capitalized? "the" makes no sense as there
are lots of large ensemble approaches. Furthermore, has already
stated above, this is not a large ensemble.

This method yields as reasonable results for an adequately resolved
parameter space as more advanced statistical techniques with means of
interpolating results between sparsely separated points in
multi-dimensional parameter space.
I would strongly dispute this since full-factorial sampling
restricts one to a relatively small number of ensemble parameters.
The cited Pollard et al (2016) paper used an ensemble of
WAIS only simulations for the last 30 kyr. Ie all ensemble
members had identical initial conditions and identical
time evolving ice boundary conditions at the junction with
the East Antarctic ice sheet. This is a far cry from applying
4 ensemble parameters to the whole AIS for 2 glacial cycles.

It covers uncertainties within the Earth model for values of
1e19, 5e19, 25e19 and 100e19 Pa s.
this study would benefit from better attention to relevant
litterature. While there is local support for viscosity as low as
1e19 Pa s on the Antarctic Peninsula, there is no support for even
5e19 over say the whole WAIS. Furthermore, the upper bound test
viscosity (and please use the more standard X10^21 units as
preferred in the GIA community) is half of the 95% "confidence"
upper bound of 2.0 X10^21 Pa s of Whitehouse et al, 2012 (GJI).

This compilation also in150
cludes records of regional sea-level change (RSL), which has not been considered
 in this study since
most of the sea-level signal is a result of the sea-level forcing with up to 140
m rather than the
model's ice dynamical response expressed in terms of sea-level equivalents, as P
ISM lacks a selfconsistent
sea-level model
Since the RSL data for Antarctica is above present-day elevation,
the above statement as written is incorrect. The RSL data is the
signal, and dominance of a far-field sea-level forcing would result
in sealevel below present-day.

Mean-square-error misfit to observed grounding line location for the modeled Ant
arc135
tic grounded mask (ice rises excluded) using a signed distance1 field
I don't understand what you mean by signed distance as RMSE would
only care about unsigned distance. Or do you mean what we do in my
group: also track mean (ie not RMSE) error and use that to assign a
signed value to the RMSE?

5. UPL: Mean-square-error model misfit to modern GPS-based uplift rates on rock
outcrops at
35 individual sites using the compilation by Whitehouse et al. (2012b, Table 2)
including
individual observational uncertainty
Would be good to update the GPS data-set. Current GPS data versus
that approaching a decade old would make a significant difference in
observations and observational uncertainties.

Then the individual score Si,j is
normalized according to the median to
Why the mean versus the median?

As in Pollard et al. (2016) we also assume that each data type is of
equal importance to the overall score, avoiding the inter-data-type
weighting used by Briggs and Tarasov (2013); Briggs et al. (2014),
which would favor data types of higher spatio-temporal density
Would you still do this if say you only had 3 ELEV or EXT
datapoints? If all data were statistically independent, then one
would demand that data types of higher spatio-temporal density would
get more weight since in this case each and every datum should have
equal weight. You need to provide a better justification for this
choice then just blind citation of previous studies.

The parameter ESIA enhances the shear-dominated ice flow and hence ice thickness
enhanced ice flow will not enhance (ie thicken) ice thickness but thin it

For the upper range of
mantle viscosities up to VISC = 1022 Pa s we find a normalized ensemble mean of
27% and 20%,
This contradicts what you previously indicated was an upper bound value of 100
X10^19 Pas on page 4. ???

This value is also used in the GIA model ICE-6G
(Peltier et al., 2015)
kind of irrelevant since Peltier doesn't do regional tuning of Viscosity profi
les.

The best score ensemble members are found for intermediate mantle viscosities of
VISC=5X10^20 Pa s and VISC=25X10^20 Pa s.
# THis again contradicts the values given on page 4. Furthermore,
# 25X10^20 Pa s is a high viscosity for the upper mantle (upper mantle is what
# this half-space model best corresponds to)

3.2 Reconstructed sea-level histories
-> ice volume histories or sea-level contribution histories

the ensemble mean ice volume is 1.0m SLE
below modern with a score-weighted standard deviation of around 2.7m SLE (volume
 of grounded
ice above 300 flotation in terms of global mean sea level equivalent as defined
in Albrecht et al. (2019).
Should compare this to published (paleo data-based) inferences of the Eemian h
igh-stand
as it makes it hard to fit current proxy-derived estimates for the Eemian high
-stand given
constraints on what Greenland could have contributed.

The LGM ice volume increases for lower PPQ, lower PREC and lower ESIA
315 values, while it seems to be rather insensitive to the choice of VISC
All these relations would be expected as such.

As MWP1a
initiated the Antarctic Cold Reversal (ACR) with about two millenia of colder su
rface temperatures,
330 a freshening of surface waters leading to a weakening of Southern Ocean over
turning, resulting in
reduced Antarctic BottomWater formation, enhanced stratification and sea-ice exp
ansion.
This is not a sentence.

The modeled range between Last Glacial Maximum and present-day ice volume by Whi
tehouse
360 et al. (2012a) is about 5.0X10^6 km3 (or 7.5 - 10.5m ESL, eustatic sea-level
 based on volume above
flotation),...
There is no point in listing all the exact ranges here and then showing those
ranges in fig 11

```
# Add the conversion factors to the figure key and have the paragraph focus on g
eneral comparison.


Briggs et al. (2014) ... from PSU simulations for 40 km resolution
# To be accurate PSU + full visco-elastic isostatic adjustment bedrock
# response with radially layered earth viscosity profile + different
# subshelf melt, basal drag, climate forcing, and calving
# treatments. So only PSU ice dynamics and thermodynamics.

Although the Large
445 Ensemble method is limited to a comparably small number of values for each p
arameter
# no it is not. A full-factorial (grid) ensemble is limited. Not other
# large ensemble approaches. And I don't understand why "Large
# Ensemble" is capitalized.  If you are choosing to equate a grid
# ensemble as a "Large Ensemble", that makes no semantic sense. How are
# readers supposed to differentiate between this useage and other
# modelling studies that will use large ensembles under a different
# sampling scheme? What about studies that have O(10) or more larger
# ensembles?

# fig 11 plot and caption: there needs to be a note that some of the indicated
# studies have non symmetry distributions. Eg, Briggs et al, 2014
# states : "likely
# between 5.6 and 14.3 m equivalent sea level (mESL), and with less
# confidence >10 mESL" = 4.0 X 10^6 km^3 of ice. The
# whitehouseBently12b datapoint is also problematic as that GJI paper
# never discusses ice volume or total sea level changes. The
# whitehouseBently12a does (and is cited in the preceeding discussion)
# and provides an uncertainty range but the plot only shows a single
# datapoint with no uncertainty range


# fig 12 please increase the font size of the colour key for those of
# use with ageing eyes...

# fig 14: please use a higher contrast colour scale, to make the
# comparison more discernable. Even better would be the addition
# of a difference plot to make clear what the differences are.

# fig 15-17 are hardly mentioned in the main text, with no consideration
# of details. Eg, figure 17 has 7 timeseries, not one of which is individually
# refered to in the text. So why is this figure in the paper?


3.4 Comparison to previous large ensemble study
# -> Comparison to Pollard et al. (2016) ensemble study
# Your current title implies there was only one large ensemble
# and the subsequent text implies it Pollard et al. (2016).
# Briggs et al was a much larger ensemble study...

while other parameters that affect the
modern grounded ice areas are sufficiently constrained by earlier studies
# This is not a fact, at best state they claim this.

In their ensemble analysis Pollard et al. (2016) included an iceberg calving par
ameter. Our 'eigencalving'
model provides a fair representation of calving front dynamics independent of th
```

e climate
435 conditions (Levermann et al., 2012).
What does "fair" mean? Be precise.

As we used the PICO model (Reese
et al., 2018) that includes physics to adequately represent melting and refreezi
ng also for colder than-
present climates, we have chosen other parameters in our ensemble,
What about the large uncertainty is subshelf ocean temperature? Above you invo
ke this
to explain the lack of any MWP1a signal compared to eg Golledge et al, 2014

comparably small mantle viscosity around VISC = 5-25X10^20 Pa s,
small compared to what? I wouldn't call these small upper Mantle viscosities f
or Antarctica

Due
to the comparably coarse resolution and the high uncertainty that comes with the
 strong non-linearity
(sensitivity) of the system we here discuss rather general patterns of ice sheet
 histories than exact
numbers.
This non-linearity is another reason to increase the number of ensemble parame
ters.

Our ensemble-mean lies at the upper range of most previous studies, except
470 for the large ensemble study by Pollard et al. (2017) with only 3-8m SLE sin
ce LGM, as their
score algorithm favored the more rigid and hence thinner ice sheet configuration
s.
incorrect in that half of your stated range is below the favoured
range of Briggs et al (2014) who state
"The LGM ice volume excess relative to present-day is likely between 5.6 and 14.
3 m
equivalent sea level (mESL), and with less confidence >10 mESL
Furthermore, your sentence contradicts itself as currently written.

Previous studies with PISM Golledge et al.
(2014) suggest
english and punctuation...

In this study we used the
Bedmap2 topography remapped to 16 km resolution without local adjustments
Does this actually belong in the "Conclusions"?

provides model and observation calibrated parameter constraints for
projections of Antarctic sea-level contributions
awkward and somewhat indecipherable. Do you just mean
"data-contrained projections of .. using PISM"?

With the best-fit simulation parameters we have
490 participated in the initMIP-Antarctica model intercomparison (Seroussi et al
., 2019, PISMPAL3).
This is not a conclusion

appendix A : is referred two a couple of times, but without any
statement of what the takeaway from the appendix is.

```
In the basal sub-ensemble we find even better scores than for the best fit param
eter combination
in the large ensemble (here no. 8102, see Fig. A.

However, best fit to the nine constraints are found for
the basal ensemble..
which agrees with the best fit values of large
ensemble.
# The above two statements contradict each other
```