Thank you once again for your comments. We are not 100% confident that we have understood your exact point. We have tried to answer your concerns in the following.

To your first concern:

There has not however been much success in answering my questions about the calibration apart from calling the ensemble member "plausible", without giving any reasons why you think this is the case. I think you might have misunderstood what I am asking. I am not asking why you do not use transient experiments for your calibration, neither am I asking for more details on what was actually done. Rather, it is necessary for you to explain the basis for the calibration experiment and what the resulting ensemble means and therefore what it is useful for. I still think that an additional paragraph to this end at the start of section 3 is required - it is confusing to jump straight from the calibration subtitle into the method subtitle without explaining why you are doing it. I am sure this stuff is inside the heads of you and your co-authors, but readers are not telepathic - I just want it written down so that readers can understand the purpose of your work!

We have modified the text of the manuscript at the start of section 3 as follow:

3 Calibration for the Antarctic ice sheet

~~3.1 Methods~~

Over the years, several GRISLI internal parameters have been shown to be of importance to appropriately simulate the flow and mass balance of the Antarctic ice-sheet. Values for these parameters have been so far derived from non-systematic tests and expert knowledge. To systematically investigate the role of those parameters and find the best fitting set for the simulated Antarctic ice-sheet with respect to the observed one, a calibration methodology with systematic exploration of the different values is performed in the following. The best fitting set will be considered as plausible models within the chosen parameter space.

~~In the following, we present a simple calibration methodology for the Antarctic ice sheet based on a large ensemble of model simulations.~~ Given its degree of complexity, GRISLI is mostly designed for multi-millenial integrations. Due to long-term diffusive response to SMB and temperature changes, an accurate methodology to select unknown parameters of the model would be to run long transient simulations with a climate forcing as close as possible from past climate states, ideally with a synchronous coupling between the ice sheet and the atmosphere. However, climate models generally fail at reproducing the regional climate changes during the last glacial-interglacial cycle as recorded by proxy data (Braconnot et al., 2012). Further- more, the phase III of the Paleoclimate Modelling Intercomparison Project (PMIP3) has highlighted the large disagreement between participating climate models in simulating the Last Glacial Maximum (LGM) in the vicinity of northern Hemisphere ice sheets (e.g. Harrison et al., 2014). Given these uncertainties amongst climate models and the large sensitivity of the ice sheet model to climate forcing fields (e.g. Charbit et al., 2007; Quiquet et al., 2012; Yan et al., 2013), it is difficult to calibrate the mechanical parameters independently from that of the SMB, in particular for northern Hemisphere ice sheets.

For these reasons, here we suggest a simple calibration methodology for the Antarctic ice sheet in which the model is run for 100 kyrs under a perpetual modern climate forcing ~~in order to reach an ice sheet~~until equilibrium.

3.1 Methods

In the following, we use the 27 km-grid atmospheric outputs, namely annual mean temperature and SMB, from the regional climate model RACMO2.3 (Van Wessem et al., 2014), averaged over the 1976-2016 time span. The basal melting rates under ice shelves are prescribed for the 18 sectors of the Antarctic ice sheet as defined in ISMIP-Antarctica project (Nowicki et al., 2016) and are shown in Fig. 3. Their values are based on the sectoral average of sub-shelf melt rates that ensured stable ice shelves (minimal Eulerian ice thickness derivative) in the recent intercomparison exercise InitMIP-Antarctica (Nowicki et al., 2016), with slight modifi- cations due to change in resolution. They are in line with observations-based estimates (Rignot et al., 2013). We do not apply any correction related to geometry changes to the climatic forcings during the calibration.

To your second concern:

I previously suggested that perhaps your calibration+transient ensemble experiment was closer to being a sensitivity test, as it shows that model versions with similar fit to the data can have different parameters and quite different results. So now I am now very confused, because your latest Author's response appears to support this line of reasoning, but the text of the manuscript does not! Maybe I have missed a paragraph in the manuscript (in which case I am sorry and please point it out to me) but I can't see where you include these points from your own response, "However, using these 24 plausible models can help the reader to grasp the GRISLI result spread (Fig. 14) for models yielding a relatively similar present-day ice sheet. We acknowledge that the choice of 12x2 ensemble members is arbitrary and this number is too low to infer statistically meaningful results in terms of model uncertainty. However, even with our relatively coarse resolution, a 400 kyr-simulation represent a nonnegligible computing time that has to be added to the 600 ensemble members of Sec. 3. As a result, Fig. 14 aims at illustrating the spread in ice volume within the 12x2 ensemble members calibrated to reproduce the present-day Antarctic ice thickness."

We have added this notion before the start of section 4 as follow:

From each of the two300 ensembles members within (AN40S and AN40T), we keep the 12 ensemble members out of 300 that have the lowest RMSE and use them in the next section for transient simulations covering the last 400 kyrs. Using these 24 plausible models on long term transient integration provides insight on the GRISLI result spread for models yielding a similar present-day ice sheet. Indeed, while they have a similar RMSE, they have distinct parameter values (Fig. 5 and Fig. 6) and as such they provide an insight of the uncertainties in ice sheet evolution relative to parameter choice. We acknowledge that the choice of 12x2 ensemble members is arbitrary and this number is too low to infer statistically meaningful results in terms of model uncertainty. Still, even with our relatively coarse resolution, 400 kyr-simulations represent a non-negligible computing time that has to be added to the 600 ensemble members of Sec. 3. We consider that this small subset contains plausible models for the Antarctic ice sheet. These models are used in the next section for transient simulations covering the last 400 kyrs. While they have a similar RMSE, they have distinct parameter values (Fig. 5 and Fig. 6) and as such they provide an insight of the uncertainties in ice sheet evolution relative to parameter choice.

4 Antarctic ice sheet changes for the last 400 kyrs

4.1 Methods

By construction, equilibrium simulations such as the ones shown in Sec. 3 do not allow for the validation of the dynamical response of the flux at the grounding line since there are no climatic transitions and subsequent migration of the grounding line. The main objective of this section is thus to show the ability of the model to reproduce large ice sheet geometry changes in response to Quaternary climate change. As a consequence of our limited knowledge of past climatic conditions in the Antarctic ice sheet region over glacial-interglacial cycles, we use here an idealised reconstruction of SMB, near surface air temperature and oceanic basal melting rates based on a limited number of proxy records. Our approach is somewhat similar to previous works (e.g. Ritz et al., 2001; Huybrechts, 2002; Pollard and DeConto, 2009; Greve et al., 2011; Golledge et al., 2014).