Reviewer response to the manuscript: 'Statistical emulation of a perturbed basal melt ensemble of an ice sheet model to better quantify Antarctic sea level rise uncertainties'

Anonymous

September 4, 2020

Based on a large ensemble of the state-of-the-art Community Ice Sheet Model simulations the authors develop a flexible and efficient method to represent the Antarctic ice sheet under a wide range of ocean melt forcings, a major part of the general climate forcing in Antarctica. The manuscript explores the use of an statistical emulator for updated projections of 200-year sea level rise contributions. The link between global climate predictions and the Antarctic response is a key challenge the ice sheet modelling community is currently facing and the use of advanced uncertainty quantification methods, as done here, is likely to play an important role for this.

The presented work is therefore of very high relevance, however, I have concerns regarding the proposed approach to find probabilistic projections and a lack of testing of this approach. These are largely evolving around refining the parameter priors which should make it easier to address my concerns for revision. Once this is done I see the potential for follow up studies to combine multi-GCM southern ocean climate prediction distributions with ocean melt prediction PDFs and the ice sheet model sea level emulator developed here. This could be an major step towards a quantification of the sea level rise projection uncertainties. I will address general concerns below, followed by specific comments.

General Comments

The main contribution of this work is to develop a framework to propagate uncertainties in ocean melt forcing (originating from a large range of sources) to uncertainties in the Antarctic mass loss. I am glad to see that this focus is highlighted by the authors throughout the manuscript. For developing such framework, however, more analysis is needed, in particular regarding the limitations of the currently proposed approach. This should include, as far as possible, a benchmark of the approach to successfully represent uncertainties.

For the currently proposed approach I am concerned about the influence of the input space boundaries on the analysis and the unclear (to me) measures taken to elevate this. Several of the optimal fitting parameters for t0 and (maybe) tau appear to be at the edge or even outside of the considered input space. As the authors note, this is a problem because large parts of derived priors lie outside of the considered space which cannot be (well) represented in the following analysis. The authors shift the fitted parameter values away from these boundaries (to a value 'within 1 sigma of best fit') which is, however, no appropriate measure to handle this limitation. First, it is not clear what data the sigma (=standard deviation?) is based on and which/how many parameter values are effected. But more importantly, by (seemingly arbitrarily) changing the values you are not representing the ocean model melt forcing anymore which undermines the following analysis.

The preferred solution would be to run additional CISM simulations to expand the input space. In case they exist, you could also give strong arguments why such simulations would have to be considered unrealistic (a-priory) to justify the prior cut-off. Otherwise you would, at the very least, have to show the impact of said boundary cut-offs on the sea level rise distribution (see quality assessment below).

The second concern are the sigmoid parameter distributions themselves (Figure 10). In some cases you have as little as two values (Peninsula) to fit a Gaussian distribution to, which leads without doubt to a large standard error of the mean. In addition you treat all sigmoid parameters as statistically independent. Let me give an example why this is problematic: The sigmoids with [t0=100, tau=70, M200=50] and [t0=0, tau=700, M200=250] result in very similar forms, despite having very different parameter values (I realize that said parameter values are outside of the parameter space you consider, but this is just an example for the existence of ambiguities). Another example is the case where the melt rate is constant in time (M200=0) in which case the other two parameters are completely unconstrained. By an independent treatment you cannot take this kind of ambiguities into account. This raises the question of how the proposed approach to find priors compares to approaches which do not rely on such an assumption of independence. These kinds of limitations are clearly stated by the authors (e.g. line 341-347), but should be addressed by further analysis (i.e. comparing and testing alternative approaches).

Two possible alternative approaches come to mind: (1) The sigmoid parameter priors could be base on the cumulative ocean melt instead of the fitted parameter values. Or (2) the sigmoid parameter distributions could be derived jointly by agreement with the time-series themselves; One possible approach to this (but by no way the only possibility) would be to take the mean and sd. for each year of the ocean model time series for each basin (gray lines in Figure 5) and search the input space for all sigmoids which are in reasonable agreement with this. 'Reasonable agreement' can by defined with a threshold or weighting on any kind of score function (such as a simple RMSE), or possibly by ruling out all sigmoids which are outside of the mean+-3sigma interval for more then 5% of the time (Pukelsheim, Friedrich. "The three sigma rule." The American Statistician 48.2 (1994): 88-91.) and uniform (equally likely) treatment of all passing sigmoids.

In any case, we need to address the quality of the resulting distributions. One easy step of validation would be to inspect the consistency of SLR contributions from the updated emulator priors (i.e. solid lines in Figure 12) with results from the optimal sigmoids (without shifting them, i.e. the ones shown in Figure 5). If there are biases, can they be explained? How compares the spread within the emulated ensemble with the optimal? Similar tests could be done for subsets (selected or random) of CISM simulations which are treated as synthetic test melt predictions (perfect model test).

As you point out, the very limited number of ocean model melt predictions limits the ability to constrain the prior distributions. Considering the focus on designing and testing statistical methods I would suggest to distinguish between 100-year and 200-year predictions more strongly. For the 100-year period more ocean melt projections are available from the current sources, but also projects like ISMIP6, which would allow for more analysis of the priors (as described above). If appropriate, inter-GCM comparisons could be another option.

Specific Comments

L1: 'and have already begun thinning in response to increased basal melt rates' Many Antarctic ice shelves are not currently thinning

L32: Consider more appropriate reference (e.g. Chapter 4 of IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, or references therein)

L44 or elsewhere: consider Holland et al. (2019) for more context (https://doi.org/10.1038/s41561-019-0420-9)

L40: -'the forefront' +'a focus area'

L41: comma after 'those that do'

L70-74: This is a good explanation of emulators

L95: here and elsewhere: ice model -> ice sheet model

L109: Since you combine (2) and (3) of the list above, I am skeptical about 'is intended to be agnostic with respect to these assumptions'. Maybe something like: 'is intended to overcome some of these limitations'?

L110: highly -> densely

L113: 'any scientific source': can you be more precise?

L143: please specify the SMB forcing here; Is the SMB constant through the spin-up and future? Which years is it based on and how?

Figure 1: The order of tasks is a bit confusing; Task 3 is (emulator) validation and Task 5 is emulation. Can Task 5 and 6 be combined?

L146: In Seroussi et al. 2020 a 30 000 year spin-up is used. What was the reason to use 40 000 years here and how big is the difference to the ISMIP6 initial state?

Figure 2: A difference plot between the modelled and observed speeds would be helpful here, e.g. replacing the observed speed map similar to Figure 3

L154: The model is in a steady state and the observations indicate otherwise, this alone can be seen as a contradiction to 'excellent agreement between observed and modeled...', please rephrase.

L156: 'very little drift' can you give a number/order of magnitude here?

L159: I agree with the editor that the interpretation as preindustrial initial state (after nudging to modern velocities and using ocean melt forcing from 2000 onward) is not helpful, nor necessary.

L165-L173: Please clarify: Are any ocean melt predictions used here based on ECHam5 simulations? ECHam5 based ocean melt simulations in both references (Timmermann and Hellmer (2013) and Cornford et al. (2015)) seem to end in year 2100 while all melt forcings shown here are 200 years long.

Please further clarify: Are any ocean melt predictions used here based on the E1 climate scenario? And if so, which? Comparing Figure 4 in Cornford 2015 with Figure 5 in this manuscript makes me wonder why, e.g. the Filcher-Ronne ice shelf time-series for HadCM3 / E1 with BRIOS ocean model is not included as Weddell sea ocean forcing. If there has been a selection process, what is the criterion for selection? A Table with source, ocean model, climate model and climate scenario would help with a more transparent database.

Which are the four East Antarctic Ice Sheet ocean melt predictions used for Figure 5? I am just asking because Timmermann and Hellmer (2013) seem to focus on only three types of 200 year simulations (HadCM3-A1B+FESOM, HadCM3-E1+FESOM and HadCM3-A1B+BRIOS) and Cornford et al. (2015) do not include the EAIS at all. On top of this there seems to be an unfortunate similarity between the highest EAIS melt prediction in Figure 5 and one of the Cornford et al (2015) predictions for the Marie-Byrd land (Figure 4, lower right panel, HadCM3/A1B+Brios). Again, the before mentioned table could help to clarify this.

L191-193: The comparison of quasi-random Sobol sequences to pseudo-random Monte Carlo sampling is unclear. What is the difference between quasi-random and pseudo-random? What do you mean by Monte Carlo sampling? Consider here the wide definition of 'Monte Carlo methods' on one side and the popular Monte Carlo Marcov Chains on the other side (which are designed to build clusters to represent priors). This is probably also a good place to mention that the Sobol sequence has a uniform distribution though the input parameter space (unless it is already mentioned in this section somewhere?).

L199-201: 'capturing all of the sigmoidal characteristics seen in the modeled ocean melt rate projections'; This creates the impression that the parameter ranges include all optimal sigmoidal fits to ocean model melt projections. Some of the lines in Figure 5 appear to have t0 values outside of the 100-225 year range. Please clarify whether this is the case and consider this in the impact/justification discussion of input space boundaries.

L224: Is the melt anomaly also imposed at newly ungrounding locations?

L231: 'highly sample' -> 'densely sample'

Figure 5 and elsewhere: m/yr vs. m/y (Figure 11) vs. m/a vs. m a^{-1}

Figure 6 caption: maybe: '...up to the maximal melt values given in the text, representing roughly twice the maximum 'data' melt rate at year 200'

Figure 8: Define the error bars and preferably avoid calling the CISM SLR values 'True' (since it is not the true future SLR contribution). You could use e.g. 'ice sheet model vs. emulated' or 'CISM vs. CISM emulator', etc.

L249: This chapter does a good job in explaining Gaussian Process emulation to non-expert audience. However, it is lacking one or two sentences with technical details about the emulator to be reproducible. E.g. 'Usually the smoothness of data being fit is estimated as part of the interpolating procedure.': is this the case here? Using a marginal likelihood optimisation? What covariance function is used? Is the nugget (representing noise in the training data) set to zero since you emulate a deterministic model? What emulator sample size is used?

L252: 'have excellent accuracy', in principal emulator and ISM values can be highly correlated but not along the main diagonal, in which case the emulator would not be accurate. Maybe just rephrase to 'show a good emulator performance' or similar.

L265: 'corresponding [+rate of] mass loss (Gt/yr)'? please clarify whether this is the rate of mass loss contributing to sea level rise (i.e. from grounded areas), total loss of ice (i.e. including ice shelves) or flux across the grounding line (i.e. dynamic contribution to grounded ice mass balance, which could be balanced by SMB). The same applies to Figure 9, labels and caption. Also consider using mm sea level equivalent instead of Gt for comparability.

L268: It is good that you are very clear about this.

L275-276: Again, it is not clear how exactly this is done but if I understand it correctly I do not see any good justification for this.

L277: What is the impact of using normal distributions? The sample size is probably not large enough (yet) to support or dismiss this assumption, but you could try other shapes of distributions to investigate the sensitivity.

L280 from 'This falls within...': consider moving this and/or the remainder of the section to the discussion.

L291: Please briefly set this work into relation to Levermann et al. (2020) (LARMIP-2).

Figure 9: Restricting the x-axis to 0-200 years would make this (and other) figure look tighter. The yellow bar in the top figure starts at 0, is that correct?

Figure 9 caption: blue shading (needs to be better defined) and red curve are introduced twice (for each panel) which might be possible to be combined. (maybe something like: Figure 9. 500-member CISM ensemble during 200-year simulations with mean (red curve) and total range of simulations[?] (blue shading). Shown are the SLR contribution (top) and rate of SLR contribution[?] (bottom)...).

L292: Agreed. I hope my remarks will help to deepen the exploration.

Figure 11: I assume the differences in 'best fits' to Figure 5 come from the one sigma shifting of the parameters alone? It illustrates well that the priors you generate encompass the 'best fits' quite well, which could be mentioned in the text.

Figure 12: This figure shows predominantly the combined effect of constraining the ocean melt forcing and the methodological choices to derive the priors, compared to the training ensemble. As mentioned in the general remarks it should also include the results of the optimal parameter fit simulations/emulations to illustrate the impact of methodological choices (on the priors) alone, as well as new SLR distributions based on additional prior(s).

L299: 'we can use it to [+densely] sample'

L309: 'the [+emulated] CISM model projects' (or, in case these numbers are not based on the emulator, make this more clear and compare to respective emulated values)

L460: The final version of this reference has become available

L485: The final version of this reference has become available