Author response for "Quantifying the potential future contribution to global mean sea level from the Filchner-Ronne basin, Antarctica"

Emily A. Hill[1,2], Sebastian H. R. Rosier[2], G. Hilmar Gudmundsson[2], and Matthew Collins[1]
[1]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom
[2]Department of Geography and Environmental Sciences, University of Northumbria, Newcastle-upon-Tyne, United Kingdom

Dear Nicolas Jourdain,

We thank you for inviting us to resubmit our manuscript after minor revisions. We have addressed the comments made by the reviewers in the manuscript, and below are our responses (in blue) to the referee comments (in black). We note that, in our previous author response document, the change to line 530 suggested by Referee 1 had not been made to the track changes manuscript. This mistake has now been rectified in this revised submission.

Best wishes,

Emily Hill and co-authors

Response to comments from Anonymous Referee 1

This manuscript describes a set of experiments using the ice flow model Úa, aimed at characterizing the spread of possible future sea level contribution from the Filchner-Ronne (FR) drainage basin. The authors create a set of ensemble simulations, under various RPC emission scenarios, that extend to the year 2300, and use these simulations to build a surrogate model. To create their ensemble, the authors sample parameters related to ocean forcing, atmospheric forcing, and ice dynamics within bounds they derive from literature, climate model ensembles of future projections, and Bayesian analysis (for the ocean forcing parameters). The authors illustrate that the surrogate model exhibits skill in predicting the sea level contribution projected by Úa in the FR region by year 2300, and they then use the surrogate to create a million-sample distribution to represent the possible spread of sea level contribution from FR within the same period. Surrogates are also created to derive estimates of sea-level contribution at years 2100 and 2200, for analysis of projected sea level contribution through time. Results suggest that the FR region is not likely to contribute positively to sea level contribution in the future, largely due to the modeled increase in accumulation over time in response to warming atmospheric conditions. However, significant contribution to sea level from this region is found to be possible (more than >30cm by year 2300), though this outcome is not probable. The author's analysis also allowed them to isolate contribution to uncertainty from the various model parameters sampled, and they find that atmospheric and ocean forcing account for the majority of the sea-level projection uncertainty, in agreement with past studies. The authors specifically highlight these model boundary conditions as the most important to improve models of in order to increase confidence in ice sheet model projections of sea level contribution.

Here, the authors present a novel approach to the challenge of quantifying uncertainties in ice sheet model projections. Running the number of model simulations required for robust assessment of uncertainties is, in many cases, not computationally feasible, so the design of an adequate surrogate model for this purpose is highly advantageous. Overall, I find that the authors thoroughly describe their methods, experiments, reasoning, and caveats. The discussion, in particularly, highlights the care that must be taken when considering results from a single ice sheet model where specific assumptions are necessary to produce realistic model ensembles. The manuscript is well-written and the figures are highly illustrative of the methods and results. The workflow diagram (Fig. 2), is especially helpful in describing the investigation's strategy. In addition, the results are thorough and well-organized, therefore I find this manuscript highly appropriate for publication in the Cryosphere, with revisions and some supporting analysis.

We are very grateful to Referee 1 for taking the time to review our manuscript and for their positive feedback.

I have a number of questions and comments for the authors, as listed below, for author response and discussion.

General comments:

Discussion section – The discussion is quite thorough, and you cover many important points and caveats. However, I think it would be improved if you also expand upon some interesting topics that are brought up in the results section, specifically pertaining to the advantages and disadvantages of using a surrogate model for this analysis. For example, in the results section, you note an example of an advantage of using the surrogate, can you expand upon this in the discussion to talk about what you learned in that exercise with respect to the importance of including extremes in your training set? Could you also expand upon what might be the disadvantages or pitfalls that others using your methods could encounter? (For example, is there a danger of not capturing threshold behavior or runaway retreat, as you observe in some of your extreme forcing simulations?) Is it possible that runaway retreat is more likely than your training set suggests, or do you think that your sampling space and final pdf capture the spread of possible scenarios accurately?
Thank you for your suggestion to include more details on the advantages and disadvantages of using the surrogate model. We have partly responded to this in the next point, and made changes to the Results section to reflect how our training ensemble and surrogate model are different/similar in terms of the projections. We have also added a couple of extra sentences to the start of the final paragraph of the Discussion to tie this through and make it clear that our surrogate model has been a powerful

tool for gaining addition insight into the future of this region, beyond what we would have gained from our (already large) ensemble alone. We had already tried to make an effort to discuss some of the disadvantages of our method in relation to capturing the regime shift seen in the Helmer et al. 2012 paper in the 6th paragraph of the Discussion. However, this was mainly due to the choice of melt parameterisation and slow increase in melt rates rather than a sudden flush of warm water that could be achieved with an ocean circulation model. We agree that this could partly be due to parameter space we have chosen for the ocean forcing parameters. One of your later comments also touches upon this, that our Bayesian inferred melt rates may be lower (but potentially more realistic), which has prevented us from sampling high melt distributions akin to a tipping point in the Weddell Sea. We've added a sentence to the 6th paragraph of the Discussion to reflect this. It is difficult to make more conclusive statements about whether runaway retreat is more likely than our training set suggests. The approach we have used does of course mean that our final pdfs are somewhat governed by the uncertainties/prior probability distributions we use for our input parameters. While we have tried to make our prior pdfs as informed as possible, which is challenging for some parameters, and in those cases we have had no choice but to assume equal probability and sample from uniform distributions. This would then effect the likelihood of extreme events and in that respect could be considered a limitation of our method. However, we have made it clear in the Discussion that our sampling and surrogate model has been able to capture a similar magnitude of basal melting to Hellmer et al. 2012, which reflects high-melt scenarios. It is of course possible that the likelihood of these extreme events could be more or less with a change in the priors, as mentioned above. Future improved constraints on our input parameters would help to provide more informed projections.

In addition, how important was it to use a surrogate to capture the full sample space? That is, do your final pdf's reveal a different pdf than your ensembles suggest? Perhaps you can show some training run (ensemble) pdf's vs. the surrogate sampling pdf's in the appendix to illustrate this point.

The final pdf's of our surrogate model versus those generated by the training ensemble are shown in Figure 6, where the tail of the surrogate model pdf extends beyond the spread of the ensemble in 2300. We appreciate that this is not easy to see for scenarios other than RCP 8.5 but are not sure we need an additional figure. The likely range (5-95%) of our results to not change dramatically between the training ensemble and the surrogate model. Despite the similarities between our ensemble/surrogate model in the likely range, we have still gained valuable additional insights by using a surrogate model. While the high magnitude contributions to sea level rise in RCP 8.5 are unlikely, they still highlight their potential within our parameter space, and what this means for ice loss, and runaway grounding line retreat. This means that our surrogate model highlighted the possibility of more extreme sea level rise scenarios that were not exposed by the original sample. We have added a further sentence to the end of Section 4.1 of the Results that makes this point point clear.

Specific questions and suggestions:

Line 1: This opening sentence is a bit awkward to read. Maybe adding "The future . . . " => "change", "behavior", or "evolution" or a similar phrase would make your point clearer.

This sentence has been revised in the manuscript to read better.

Line 34: . . . as the "combined area" of the two major drainage basins. . . (or something similar)

This has been added in the manuscript

Line 55: I understand the point you are trying to make in this sentence, but it reads awkwardly. Please try rephrasing.

Agreed. This sentence has been rephrased in the manuscript

Line 159: Could you add a thin or dashed line to Fig 1 or its small inset that shows where the divide between the two basins sits?

A dashed line and labels for the two drainage basins has been added to the inset of Figure 1, and the caption updated to reflect these changes.

Line 162: Perhaps in the supplement, an illustration of what your mesh looks like, perhaps in some key locations, would be very helpful.

A Figure of the model mesh has been added to the supplement, which includes an inset that shows the mesh around the Rutford Ice Stream in more detail. A reference to this figure has been added in the main text.

Line 170: Could you please specify the settings for your initial mesh adaption, for instance: What is your minimum mesh for the adaption? Is it still 900m? Is there a maximum mesh size near the grounding line? How close to the grounding line is the adaption imposed (i.e. is there a set buffer length)?

Additional details on the element sizes have been added to the text, including that elements within a 10 km distance of the grounding line are 3 km and 900 m within a 1.5 km's of the grounding line.

Line 171: What is the minimum thickness value that is imposed?

30 m. This has been added to the text.

Like 235: You discuss only changes to surface accumulation through time. Do your simulations have a representation of surface melt as well (i.e., a PDD scheme or something similar)? Please specify this in the text.

We have now added a sentence to the text in Section 3.3 that makes it clear that we are not representing surface melt using a PDD scheme for these regional simulations (see response to Referee 2 on a similar point). We have also added in some justification, as the paper suggested by Referee 2 (Kittel et al., 2021) does not suggest that this region will experience negative surface mass balance, even under RCP 8.5 forcing.

Line 315: The term “point” is used a number of times in the text to refer to a location in your sampling space. This terminology is easy to confuse with a point in map space. Is it possible to use a more specific term, like “sample point” or something similar throughout the manuscript to distinguish between sampling space and actual 2d map space?
We have been through the manuscript and added sample point to make it clearer throughout that this is a point from a parameter space.

Line 322: Please quantify the model drift (or spread of model drift for all of your control runs) here.
We have now added in the range of model drift in mm of ΔGMSL for the control runs

Line 328: Throughout the manuscript, you refer to this set of simulations as your “ensemble”. It would be helpful for the reader if you explicitly call out here that these are the runs that you will hereafter call the “ensemble” (or create a name for this set of runs that you can refer to later in the manuscript).
We have added 'which we hereafter refer to as our "training ensemble"', where it is first mentioned in the manuscript (the caption to Figure 2) and have changed it throughout the manuscript to make it clear which ensemble we are referring to.

Line 339: Please add a quantitative statement with respect to the surrogate model being “close” to the ice flow model response.
We have now calculated the root mean square error between our modeled responses (from our validation sample) and the predictions from the surrogate model using the same validation X sample. Some lines have been added to the text, and RMSE for each surrogate model included on Supplementary Figure S2

Line 345: What type of algorithm is used for the sampling of these 1 million simulations? Is it Latin hypercube as in the other ensemble sampling?
Yes this was done using Latin hypercube sampling, and that has been added to the text.

Figure 5: Please note the year at which the sea GMSL represents in the plot or caption (i.e. 2300).
The caption already includes "by the year 2300"

Line 440: Please note in the text the simulations used for this analysis (i.e., the ensemble)
This has been added.

Line 530: While Ritz et al., 2015 do not change surface mass balance through time, Schlegel et al., 2018 apply a step function on accumulation, so there is still a possibility for suppression due to accumulation. The difference is more likely due to your treatment of ocean forcing (e.g., PICO with Bayesian exploration extreme melt rates, which may be lower but considered more realistic, especially with consideration to the possible time lag between atmosphere and ocean warming), as well as your application of no melt on partially floating elements, adaptive grounding line mesh, and even the repetition of inversion procedures for each model simulation.

This is a good point that there is a distinction between these two studies that we hadn't made clear in the text. We've separated the two, and brought the Schelgel reference and some additional explanation into the previous sentence. We have made reference to our Bayesian inferred melt rates in a following paragraph of the Discussion when we compare our results to those of Hellmer et al. 2012.

Line 535: Your results show a strong dependence on accumulation, and the discussion below gives an inclusive overview of the challenges in forcing accumulation on ice sheet models in general. Could you offer a quantitative comparison between the spread (pdf) of the change (anomaly) in precipitation that your sampling imposes on your simulations, and that that is predicted by CMIP5 models? For example, you present the spread in the ocean forcing time delay parameter from LARMIP-2, could you show similarly, maybe in the supplement, a similar pdf for your regional accumulation from various CMIP GCM's and then compare that against a pdf that represents accumulation sampling from the surrogate (or even just the ensemble?). I am curious to see this comparison, since while you sample p in a normal way (Fig. 3), this parameter is exponentially related to accumulation. Is it possible that this choice skews your sampling to higher sensitivity to temperature change? Is the imposed sampling of total increased accumulation realistic as compared to what GCM's are predicting? If not, this should be mentioned as a caveat in your discussion, because it has implications for the shape of your final GMSL pdfs (Fig. 5), and total probabilities.

This is a really interesting and relevant point. It is still an open question in the field as to whether parameterisations of accumulation change with warming used in ice sheet models are able to accurately replicate accumulation predictions in CMIP GCMs. It would almost be a separate study in itself to validate these, including selecting appropriate CMIP5 models to use, and extracting this information regionally for the FR basin. It would make for interesting future work, but is perhaps beyond the scope of this study. In the case of the LARMIP-2 time delay parameter, we were directly using this information as a prior to our Bayesian inference of melt rates. Of course if we were to tune the input distribution of our p parameter, based on it's ability to produce accumulation rates similar to GCMs then it would be appropriate to use the output of CMIP models as a target to constrain the distribution of p. However, doing this would also need some careful consideration as CMIP models themselves show large temporal and spatial variability in their predictions of precipitation, and some of them may not accurately replicate current precipitation changes. The values for p we use are taken from studies that have performed such an analysis and selected CMIP5 models that are able to replicate observations to derive the value of p (e.g. Palerme et al., 2017 - Climate Dynamics). The work of Rodehacke et al. 2019 also showed some of the challenges and stark contrasts between projections of sea level rise using either CMIP5 results directly, or a precipitation scaling, and perhaps suggests that using such scalings is a better option than CMIP results, due to the uncertainties in CMIP data themselves. Using the spatially variable scaling they propose, would be an improvement to our approach in future. We can agree that sampling from a uniform distribution may have skewed the distribution of accumulation and had an impact on our on projections of ΔGMSL. In the absence of additional information on the distribution of p (that indeed in the future could be constrained by observations/CMIP results), it is difficult to know what other distribution we could have used. Our approach here was really to say "this is the state of current parameterisations and knowledge of uncertainties in some of the parameters that are used and what happens when we propagate these uncertainties through our model" and not to validate the results to other model results. But we can agree that this is indeed a caveat to our results, and we have added a couple of sentences to the Discussion to reflect this.

Line 764: Could you quantify at what probability does the Reese et al., 2018a value sit in your distribution? Can you comment on the possible implications of your values sitting much lower in magnitude than suggested by this previous study? In Fig. S2 it appears that once the γ_T^* value approaches 2x10^{-5}, that the delta GMSL contribution starts to become highly negative. Can you discuss why that is and how it influences your resulting pdf from Fig. B1?
We have gone back and calculated the percentile in our distribution that the Reese et al., 2018 value sits, and it is indeed at the very extreme of our distribution at the approx. 99th percentile. It is not too surprising that we found a lower value, given that their goal was to calibrate an appropriate value for both Pine Island and Filchner-Ronne. It is a good point that the results of Figure S2 suggest that the GMSL contribution becomes highly negative in this region. I think it's important to note that the pdf in Figure B1 is only used as an input to our main uncertainty analysis, and is therefore not affected by the GMSL projections or main ensemble of simulations. However, it does of course affect the uncertainty range of our priors, and subsequent sampling to generate our training ensemble, and additionally the 1,000,000 point sample that was used to get our projections from the surrogate models. It is probably the case that because there were few training samples beyond 2x10^{-5}, that the surrogate model ultimately struggles to replicate the individual response to this parameter alone for these high values.

Line 766: basal mass balance?
Changed.

Figure S1: Please specify the simulation year of GMSL presented.
"By the year 2300" has been added to the figure caption

Figure S2: The results presented here are a bit puzzling. Are they, as other results, represented as difference from the control simulation? If so, if I am understanding your methods correctly, then I would expect all graphs to cross a value of zero ΔGMSL, when the value of the sampled parameter is equal to the value of the control run. Could you speak to this, specifically why most of the graphs show a negative contribution to sea level for all of the sampled values?
This is a good point, and it is indeed correct that if the only difference between our control run and perturbed RCP runs was

the change in each parameter value, then they would indeed cancel out and cross zero. However, our control runs also do not contain time dependent forcing, whereas our perturbed runs are applying a temperature change that varies through time, which is why this is not the case. We already state that the control run has constant temperature forcing in Section 3.5, so do not feel that any further changes are needed to the text.

Figure S4: Please check the lettered labels on the right-hand column plots.
Lettered labels have been corrected

Response to comments from Nicolas Jourdain

One of the two nominated referees had to cancel his/her participation for personal reasons. I nonetheless received a positive preliminary feedback on this work. To save time in the review process, I have therefore decided to write a review myself.
Thanks to Nicolas for stepping in to provide a second review of the paper at the last minute, we appreciate the additional feedback provided.

The paper reads well and the methods are robust and clearly described. The results are important for the ice-sheet and sea-level communities, and I only have a few minor comments that will hopefully improve the paper:
Thank you for the positive feedback on our manuscript and for your suggested improvements. As before, we respond to the comments below in blue text.

L. 136: effect -> effects
Changed

L. 165: "with a percentage deviation of only 3%" -> specify in 2100 or 2300.
Added this is for 2300 and made it clear it is a deviation in ΔGMSL

Fig. 3: indicate the units of tuned coefficients.
Units have been added for C and γ_T^* and the time delay. We've also taken the opportunity to update the value of p on the x-axis as a percentage rather than a fraction for clarity.

Section 3.3: the increased surface mass balance for higher temperatures holds for moderate warming, but for RCP8.5 warming to 2300, there will likely be more ablation by surface melting and the surface mass balance may become negative at some locations (see Kittel et al. 2021, their Fig. 5 where the negative runoff contribution of the grounded-ice SMB starts to significantly increase towards 2100). This is likely somewhat captured by the lower bound of the p parameter which is well below what is expected from the Clausius-Clapeyron relationship, but this could be briefly discussed.
This is an important point and we are grateful for the paper recommendation. Following a comment from Referee 1 we have now made it clear that we have not used a PDD melt model, and are therefore not imposing any surface melt. We have also added a sentence to the end of Section 3.3 that states that the lower bound of the p parameter is capable of capturing low rates of total surface mass balance that would occur if there was surface melting under RCP 8.5 forcing, although Kittel et al. 2021 suggest that an increase in surface runoff/negative smb is unlikely to be widespread across the Filchner-Ronne ice shelf in comparison to other regions.

L. 215-221: please provide references, e.g. to IPCC-AR5, and indicate whether the provided warming values refer to the CMIP5 multi-model mean or to the MAGICC emulator.

We have added a reference to the IPCC report and made it clear that these are the multi-model mean estimates presented in the AR5 report.

L. 275: a better or additional reference here would be Favier et al. (2019) in which the box model ("PICO") was evaluated as a relatively good parameterization: Favier et al. (2019). Assessment of sub-shelf melting parameterisations using the ocean–ice-sheet coupled model NEMO (v3. 6)–Elmer/Ice (v8. 3). Geosci. Mod. Dev., 12(6), 2255-2283.

Thank you for suggesting this paper, we have added it to the text.

L. 280: indicate that this is sea-floor temperature and salinity.

Added

Discussion section: There is a deep uncertainty related to processes that are not represented, e.g. evolution of the position of the calving front, hydrofracturing due to higher surface melt in the future, evolution of ice damage. Some of these processes may play a key role in the future FRIS contribution to sea level, and this should be discussed.

We have added a couple of sentences to the final paragraph of the discussion that highlights these key areas of additional uncertainty in projections of sea level contribution.

Response to additional external feedback

We are grateful for this additional external feedback that raises a number of important points for clarification. Again our responses are below in blue

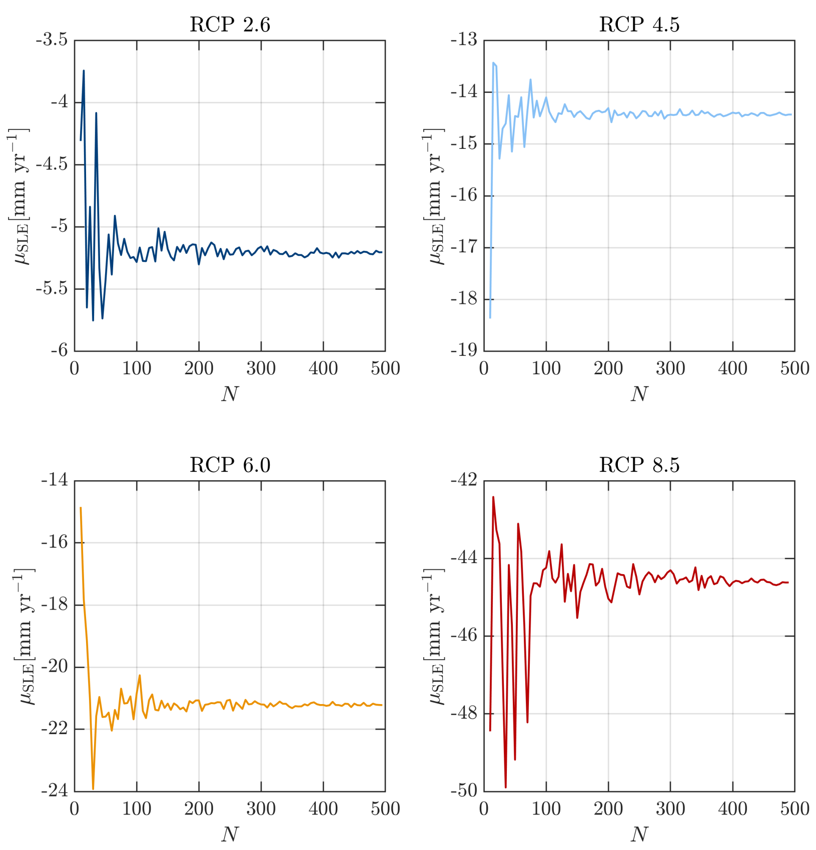
- A single model realisation is used to create the 2300 projections. It would have been more rigorous to use several simulations, in order to test the extremes of plausible scenarios.

While this is true, we do obtain a single climate realisation from the MAGGIC simulations between 2100 and 2300 for each RCP scenario, we are still capturing a spread around these projections by imposing error bounds on global mean temperature change (shaded regions in Figure 4). Therefore it is almost as if we have used an ensemble of simulations, which is captured in our parameter T, that samples a climate realisation from within the shaded uncertainty regions, and is certainly better than using a single projection.

- A Latin hypercube is used to sample 500 parameter values. Why was 500 chosen as the design size ? Has anything been done to check the plausibility of the values or to sample extremes?

We considered a 500 member ensemble to be a sufficiently large ensemble, and beyond that it becomes computationally challenging. We have also tested that the mean and variance of our surrogate models have converged based on the number of samples used in the training set to create the surrogate models. To do this we took the original "training" ensemble, and randomly extracted members from this, in sample sizes of 10 to 500 at intervals of 5 samples, we then created a surrogate model using each of these values for N (99 additional surrogate models for each RCP scenario). This reveals that $N = 500$ (or potentially less for some RCP scenarios) is sufficient for the mean and variance of our projections to converge. We've added a sentence into the text and the figure below into the supplement.

To answer the second question, the assumption here is that any of the values within the uncertainty ranges we specify for our input parameters are plausible (see response to final point too). This is because the bounding values for these parameters are informed based on those presented in the literature, which have therefore already been determined as "plausible". It would obviously be interesting to go back and test this further, and in particular, validate our model responses based on observations before using that parameter set in a future simulation. We partly did this for our box model parameters, but agree it could be extended to other parameters. This would make interesting future work.



- Several parameters are given uniform priors without much reasoning. Why was uniform the best choice?

In the absence of better constraints from prior studies or observations, for some parameters, it is not possible for us to make a more informed choice about the priors. As a result we choose not to restrict the sampling in any way, and prescribe only an upper and lower limit, and assume equal probability (uniform priors). It is a whole exercise in itself to determine more informed priors for these parameters based on available data and observations. We did choose to do this for the parameters used in the basal melt parametersation, because we were able to do this efficiently with the box model, and we had good observations in order to constrain the parameters. However, for other parameters, in particular those related to the inversion, determining the values most likely to replicate observations, is still an open question and would require a separate study. In response to a comment by Reviewer 1, we have added a couple of sentences to the discussion that mention uniform priors for p as a caveat.

- The long tail for RCP8.5 contributions at 2300 is mentioned as being due to certain parameter combinations – what combinations? Are they plausible? That would tell a lot about how much attention should be paid to these upper values.

We could argue that any of the parameter combinations within our parameter space are plausible, given our current understanding of the bounds/uncertainties related to these input parameters. We have made the statement in the text that we believe this long tail to represent large contributions to sea level rise, but is highly unlikely, as it lies outside of the 5-95th percentiles (see Figure 6). However, in hindsight, we think that the original statement in the text may have been a bit misleading, as not only do we not include any follow up, or discussion, of the combinations, but our Sobol indices reveal that the variability in RCP 8.5 contributions is primarily due to variations in a and is not dependent on specific combinations of parameters. For clarity we have simply removed the statement "certain parameter combinations".