Point by point response to review#2

In the following we will respond to all comments of review #2, the original comments in blue, response in black and changes to the manuscript is quoted at the end of each response, where appropriate. We believe we can address all concerns in a convincing manner and think that the manuscript would greatly benefit from this revision.

In the manuscript, Wernecke et al., present a promising method to calibrate uncertainty distributions of mass loss derived from ice-sheet model simulations with spatial data. [...]
Before considering it for publication, I recommend additional analyses, a more detailed discussion of the capabilities and limitations of the method and reframing as explained in the comments below.

Major comments:
• p.1 l.9, l.11, & other: with some more analysis, this study can make a very good test case that demonstates the capabilities of the new method. However, it is problematic to say that in this study you are estimating future sea-level contribution or that you are making 'predictions' or 'projections', since your analysis is based on simulations with constant ocean forcing, excluding for example natural variability (e.g., Jenkins et al., 2016) or potential future changes in ambient oceanic and atmospheric conditions (e.g., Holland et al., 2019) depending on the different socio-economic pathways (RCP scenarios). Possible future evolution of surface mass balance is not considered and uncertainty in basal melting is based on a simple amplitude scaling, neglecting for instance the effect of changes in spatial melt rate distributions (discussed, e.g., in Goldberg et al., 2019) or uncertainties related to the basal melt rate parameterisation (see, e.g., Favier et al., 2019).

We agree that we should have been more clear about the limitations of our projections. In revision we would like to turn this study into a test case as suggested.

We are not using the word 'prediction' for the model simulations used here any more. We also note that climate scenarios are expected to have small net impact on 50 year simulations and add:

"*Relating climate scenarios to local ice shelf melt rates is associated with substantial uncertainties itself. CMIP5 climate models are inconsistent in predicting southern ocean water temperatures, so that the model choice can make a substantial (>50\%) difference in the increase of ocean melt by 2100 for the ASE \citep{naughten2018}. Melt parameterisations, linking water temperature and salinity to ice melt rates, can add variations of another 50\% in total melt rate for the same ocean conditions \citep{favier2019} and hence add another level of uncertainty. The treatment of melt on partially floating grid cells further impacts ice sheet models significantly, even for spatial resolutions as low as 300~m \citep{yu2018}. It is therefore very challenging to make robust climate scenario-dependent ice sheet model predictions. Instead we use projections of the current state of the ASE for a well defined set of assumptions for which climate forcing uncertainty is simply represented by a halving to doubling in ocean melt.*

Naughten, Kaitlin A., et al. "Future projections of Antarctic ice shelf melting based on CMIP5 scenarios." Journal of Climate 31.13 (2018): 5243-5261.

Favier, L., Jourdain, N. C., Jenkins, A., Merino, N., Durand, G., Gagliardini, O., Gillet-Chaulet, F., and Mathiot, P. (2019). Assessment of sub-shelf melting parameterisations using the ocean–ice-sheet coupled model nemo (v3. 6)–elmer/ice (v8. 3). Geoscientific Model Development, 12(6):2255–2283.

Yu, Hongju, et al. "Retreat of Thwaites Glacier, West Antarctica, over the next 100 years using various ice flow models, ice shelf melt scenarios and basal friction laws." The Cryosphere 12.12 (2018): 3861-3876.
"

We thus argue that due to the large and multi-level uncertainty in RCP forced simulations the simple ocean melt scaling can be considered a representation of climate forcing uncertainty. This is not to say that we predict the future but that we do not neglect uncertainty in the forcing altogether. As long as we are not able to robustly propagate uncertainties through every level of the mapping from climate scenarios to sub-ice shelf melt, we consider a simple perturbation approach most appropriate.
In general the study will be re-framed towards a methods test, by adding a new synthetic model test and comparisons with different calibration approaches. This further reduces focus from the SLR projections. The spatial retreat probabilities section will be removed.

• p.5 l.11: the choice of calibration of dh/dt after running the model for 7 years appears random. Please explain this. Also, how would your results be influenced if your calibration was done after 1, 5 or 10 years?

The rationale to use dh/dt fields for calibration is the following. The variety of datasets available to calibrate ice sheet models is limited. Reliable and spatially-resolved satellite observations which could be useful for calibrations are limited to surface ice velocity, surface elevation and the corresponding rates of change. The surface velocity is used for model inversion and is therefore not an independent parameter. The absolute ice thickness (equivalent to using ice surface elevation with a fixed bedrock) is also set in the model parameter inversions and in addition only semi-continuous (as it cannot become negative). This causes additional challenges as described in Chang et al. (2019). We avoid these challenges by using ice thickness change data (which can be considered fully continuous as long as changes in ice thickness are smaller than the total thickness so that negative and positive values are equally possible).

Regarding the period, we compare several calibration periods and find a short spin-up phase of three years to be beneficial. Calibrations after this spin-up on the first four years, seven years and from the fourth to the seventh year all produce very similar results with projections for the end of model period of 18.4 [10.5, 26.3], 18.4[11.7, 25.4] and 17.4 [10.9, 24.6] mm SLE (weighted mean and 5.- and 95- percentiles), respectively. In the spin-up period the model adjusts to to the boundary conditions and calibrating on this period with the proposed approach creates a tendency towards slower ice sheet model runs and an underestimation of sea level contribution. We will change the analysis accordingly in a revised manuscript.

Chang, Won, et al. "Ice Model Calibration Using Semi-continuous Spatial Data." arXiv preprint arXiv:1907.13554 (2019).

• p.12 l.3: my understanding of Nias et al. (2016) is, that inversion techniques were used to estimate the spatial fields of viscosity and basal traction coefficients. Were different inversions run for the different bed geometries and values of m? If the inversion was run only for m = 1, a better fit for m = 1 in comparison to m = 1/3 would not be a surprise as the parameter fields were optimized for this case. If this is true, your findings are maybe more due to the experimental design rather than being physically interpretable. Please clarify this (similar for the bed topography and the other parameters) and, if applicable, consider it in the discussion of your findings.

Thank you for the suggestion. However, Nias et al. (2016) used different basal traction coefficient fields for the different sliding laws and bed geometries. This has been clarified in the manuscript.

• p.14 l.24-27 and Appendix B: you state that your method improves calibration with aggregated variables. It is interesting to see the effect on the different parameters (Figure B1), but to make this point clear, please add also the effect on the mass loss and grounding line probability estimates (similar to Figures 5,6).

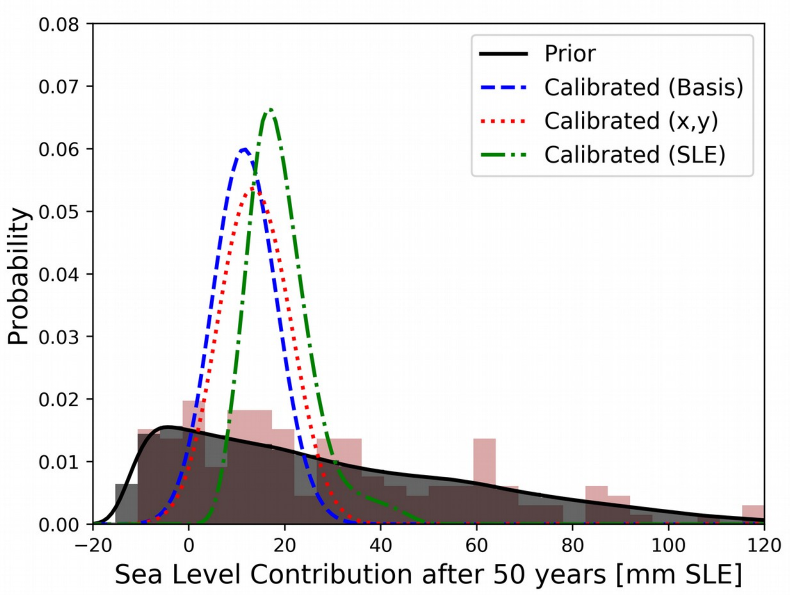
We now address the impact of different calibration approaches in more detail. This is done on a synthetic model test and for projections in a new section which is dedicated to this topic. We further compare the mass loss distributions as requested. Below are parts of this new section.

“*\subsubsection{Comparison with other calibration approaches}\label{sec:comp}*
To put the likelihood distribution from figure \ref{fig:mtest} into context, we try two other methodical choices. The first is by calibrating in the spatial domain after reprojecting from the emulator results to the principal components.

The second is to calculate the yearly sea level contribution for each set of input parameters and use this, combined with the mean observed sea level contribution for calibration.

The calibrations in basis (Figure \ref{fig:mtest}) and (x,y) representation behave very similarly, indicating that our approach is robust towards the decision to use the basis representation. Using the sea level rise contribution constrains the parameters weakly; it shares the limitations of our approach by not constraining the ocean melt and favouring linear sliding but in addition, a wide range of traction-viscosity combinations perform equally well and there is no constraint on bedrock. Furthermore, the model run used as synthetic observations is not identified as the most likely setup when the sea level rise contribution is used for calibration. This demonstrates the value of the extra information - and stronger parameter constraints - provided by the use of two-dimensional observations.“

And we added the two additional calibration approaches to the sea level rise contribution projection



Caption: Total sea level contribution from the Amundsen Sea Embayment after 50 years for $m=1/3$. The prior (black line) and calibrated (colored lines) distributions are shown based on emulation while the histograms show the prior BISICLES (red) and emulated (grey) ensembles.

Further comments:
• page 2 lines 22ff: there are a number of modelling studies with coarser resultion that do not require a parameterised grounding line for retreat (e.g., Schlegel et al., 2018). 'Regional' is maybe more appropriate than 'one glacier' (e.g. Arthern and Williams, 2017).

We now clarify that we are talking about challenges of adequate representations of the grounding line in low resolution models in general and make sure not to imply that there are no useful low resolution model studies without sub-resolution parameterisation. We also follow the suggestion of using 'regional'.

• p.2 l.28 and l.20: please check your use of 'predicted' versus 'projected'.

We do not use 'predictions' for the model simulations used in this study any more

• p.3 l.23-29: emulation of model output was also used for example in Levermann et al. (2014).
Corrected

• p.4 section 2.1: since basal melt is the driver of mass loss in the Amundsen Sea at present, more details should be given here, e.g., how do mass fluxes compare to observations?
We added:
"The ensemble covers a wide range of sea level rise contributions for the 50 year period with the most extreme members reaching -0.19 mm/year and 1.62 mm/year, respectively. About 10% of the ensemble shows an increasing volume above flotation (negative sea level contribution) with the central runs (0.5 for traction, viscosity and ocean melt parameters) contributing 0.27 mm/year (linear sliding) and 0.26 mm/year (nonlinear sliding). The average contributions are generally reasonably close to satellite observations (0.33 ± 0.05 mm/year from 2010-2013 (McMillan et al., 2014)) with 0.30 mm/year for linear sliding and modified bedrock, 0.37 mm/year for linear sliding and Bedmap-2, 0.38 mm/year for nonlinear sliding and modified bedrock and 0.51 mm/year for nonlinear sliding and Bedmap-2 (Nias et al., 2016)."

• p.5 l.13: you could state here that your $y(\theta_i)$ is dhdt.
Done

• p.5 l.16: $\Theta = [0, 1]^5 \subseteq \mathbb{R}^d$?
Clarified

• p.5 l.21: shouldn't $S \in \mathbb{R}^{m\times n}$, $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, since U, V are unitary matrices and by definition quadratic? Please check also the other matrix dimensions.
You are right, we got sidetracked by S being diagonal but not square. Thank you.

• Section 3.1: a reference to Fig. 1 is missing.
Added

• Figure 1: please give here more explanation, e.g., of 'unit length'.

Replaced ‘unit length’ by orthonormal and added ‘representing the main modes of variation in the model ensemble’

• p.6, l.8: would it be an option to calibrate not only after 7 years but at all datasets from Konrad et al. (2017) individually as they find variations in the onset and propagation of surface lowering?

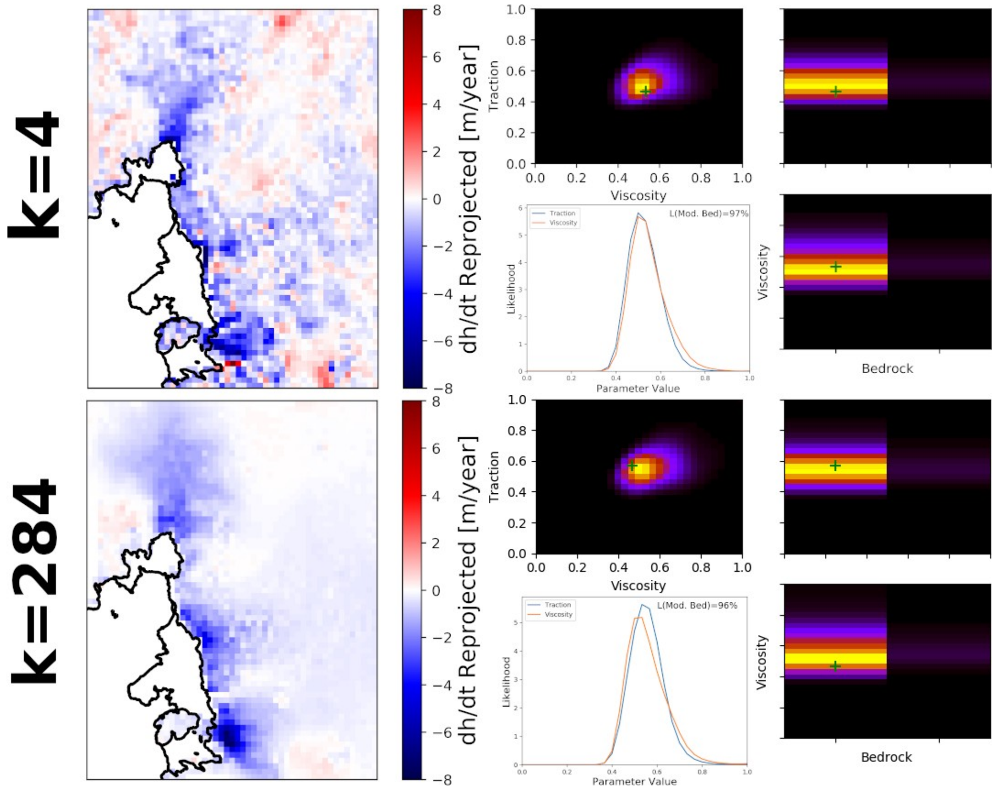
A spatio-temporal calibration would be a logical next step and is now mentioned in the discussion, but be believe this would exceed the scope of this study.

• Figure 2: in your reprojection of the mean observation, artifacts of thickening occur. How will this affect your calibration?

• p.7 l.1: a value of 0.6 seems to be rather large, please explain.

Combined:

By increasing the truncation value k we can investigate how said artifacts influence the calibration. At the same time, the fraction of the observations which cannot be represented by k principal components, as evaluated by the remaining spatial variance, diminishes. When all PCs are used (k=284) this value reduces from 0.6 to 0.045 and the thickening artifacts mostly disappear (see figure below). At the same time the likelihood distribution does change only marginally, therefore the affects of both these factors is small.



Caption: Re-projected mean observations (left) and likelihood distribution (right) for truncation value k=4 (top) and k=284 (bottom).

• p.7 l. 5: I cannot find where this is discussed in the results section?

It was not discussed but the BISICLES ensemble runs are now added as histogram in the SLE distribution plot (see above) to illustrate the improved representation by using emulation. This is now also mentioned in the text.

• p.7 l.7: you could help the reader if you explain what the rows of S 0 T 0T represent.

Done:
"*A row of S 0 V 0 T can be understood as indices of how much of a particular principal component is present in every ice sheet model simulation.*"

• p.7 l.7: how is the training done? please give more details here.
• p.7 l.12: I cannot find the definition of a Gaussian Process Emulator in the given reference.
• p.7 l.15ff: more details are needed here.
Combined:
We now additionally feature Equation 2.19 from Rasmussen and Williams (2006) in the manuscript which describes in detail how the emulator predictions are based on the training data and hence how to understand the training process. In this context more details are also added to the description of the covariance function and how exactly it is used. We also reference the python functions which are used for training and marginal likelihood optimization.

• p.8 l.16: eqn.3
corrected

• Section 3.4: you are switching between observational errors and model errors in this section. It might be easier to read if you give and explain one by one.
Has been rearranged

• p.10 l.11: prediction, see above
Corrected

• p.15 l. 28: 'the' too much
Corrected

• p.16 l. 4: please specify 'uniform within the parameter space'.
Rephrased:
"*The emulator performance, as described above, shows no dependence on the input parameters*"

• Figure A2: how are the quantities shown on the x and y axis obtained?
We expanded the description and added the mathematical nomenclature used elsewhere.

• Appendix B: It would be great to see also how your method compares to calibrations using a spatially aggregated, temporal evolution of mass loss as used for example for targeted parameter optimization in Golledge et al. (2019).
We increased the use of spatially aggregated quantities to compare the calibrations but think that a temporal calibration would exceed the scope of this manuscript.

References

Arthern, R. J. and Williams, C. R. (2017). The sensitivity of west antarctica to the submarine melting feedback. Geophysical Research Letters, 44(5):2352–2359.

Favier, L., Jourdain, N. C., Jenkins, A., Merino, N., Durand, G., Gagliardini, O., Gillet-Chaulet, F., and Mathiot, P. (2019). Assessment of sub-shelf melting parameterisations using the ocean–ice-sheet coupled model nemo (v3. 6)–elmer/ice (v8. 3). Geoscientific Model Development, 12(6):2255–2283.

Goldberg, D., Gourmelen, N., Kimura, S., Millan, R., and Snow, K. (2019). How accurately should we model ice shelf melt rates? Geophysical Research Letters, 46(1):189–199.

Golledge, N. R., Keller, E. D., Gomez, N., Naughten, K. A., Bernales, J., Trusel, L. D., and Edwards, T. L. (2019). Global environmental consequences of twenty-first-century ice-sheet melt. Nature, 566(7742):65.

Holland, P. R., Bracegirdle, T. J., Dutrieux, P., Jenkins, A., and Steig, E. J. (2019). Climate forcing of the west antarctic ice sheet: Anthropogenic trends and internal climate variability. Nature Geoscience.

Jenkins, A., Dutrieux, P., Jacobs, S., Steig, E. J., Gudmundsson, G. H., Smith, J., and Heywood, K. J. (2016). Decadal ocean forcing and antarctic ice sheet response: Lessons from the amundsen sea. Oceanography, 29(4):106–117.

Konrad, H., Gilbert, L., Cornford, S. L., Payne, A., Hogg, A., Muir, A., and Shepherd, A. (2017). Uneven onset and pace of ice-dynamical imbalance in the amundsen sea embayment, west antarctica. Geophysical Research Letters, 44(2):910–918.

Levermann, A., Winkelmann, R., Nowicki, S., Fastook, J. L., Frieler, K., Greve, R., Hellmer, H. H., Martin, M. A., Meinshausen, M., Mengel, M., et al. (2014). Projecting antarctic ice discharge using response functions from searise ice-sheet models. Earth System Dynamics, 5(2):271–293.

Nias, I. J., Cornford, S. L., and Payne, A. J. (2016). Contrasting the modelled sensitivity of the amundsen sea embayment ice streams. Journal of Glaciology, 62(233):552–562.

Schlegel, N.-J., Seroussi, H., Schodlok, M. P., Larour, E. Y., Boening, C., Limonadi, D., Watkins, M. M., Morlighem, M., and Broeke, M. R. (2018). Exploration of antarctic ice sheet 100-year contribution to sea level rise and associated model uncertainties using the issm framework. The Cryosphere, 12(11):3511–3534.

All new references have been added to the manuscript