# Point by Point response to review 3

Andreas Wernecke for all authors of the manuscript

Review of "Spatial probabilistic calibration of a high-resolution Amundsen Sea Embayment ice-sheet model with satellite altimeter data " by Wernecke et al.

I was not a reviewer of the previous versions of the manuscript.

This paper presents a dimension-reduced approach to calibrate model projections against observations of surface elevation rates of change. Following the comments of Reviewer 1, the new version includes new benchmark simulations to show that the method allows to recover correct known parameters values. The method is clearly described and the applications convincing, providing a valuable contribution to the field.

We would like to thank you for the thoughtful comments, thanks to which we have been able to increase clarity on several aspects of the manuscript, including the calibration data, the initialization process and its interaction with calibrations.

However, I still have few major comments that the authors should consider before publication.

Major comments :

- Following comments from Reviewer 2, the distinction between the simulations presented in the paper (using constant forcing) and " projections " is still unclear. The word " projections " is still used in several places to describe the simulations and this really needs to be clarified. I suggest to avoid the term " projection " in the abstract, discussion and conclusion. I suggest to split the section "2.1 Ice sheet model ensemble" in two subsections, the first to describe the model initialisation and the set of perturbed parameters, the second to describe the transient simulations, with the spin-up, calibration and forecasts periods. This would be the good place to discuss the assumptions in the forecast period and why the results differ from " projections ".

  We followed your suggestions and do not use 'projections' to refer to the simulations used here throughout the manuscript. The term "projection" is still used in the Abstract and Conclusion (once each) when it is clear that it is not referring to the simulations used here but instead to highlight the potential value for future applications. We also split the Model data chapter, as suggested. The second subsection includes the following new paragraph:

  *"The simulations used here are not intended to be predictions of the future but instead project the current state of the ASE glacial system with a constant recent-past climate forcing and perturbed parameters into the future. No change in the climate is represented in the ensemble. End-of-simulation sea level contribution distributions are presented to illustrate*

*and compare the value of calibrations and should not be understood as best estimates of future sea level contribution."*
We hope this highlights the limitations of the ensemble in very clear terms. We decided to maintain the discussion of climate forcing uncertainty as we feel the need to highlight that there is no robust link between climate change and decadal ocean melt forcing in the ASE. Therefore we think that the constant forcing is no compelling reason to discard these simulations as toy example/unrealistic.

- It is not exactly clear which observations are used and what is their equivalent in the model. The dataset is a compilation of surface elevation changes from 1992 to 2015, but we understand that only observations from a 7 years period are used. Which period? How is it chosen? Does the initialisation of the model correspond to a given date? What exactly are the model outputs that are used for the comparison with the observation, i.e. the mean surface elevation change during the 7-year calibration period, the surface elevation change ate the end of the calibration period or the average of the annual (or bi-annual) surface elevation changes during the calibration period? By the way, the term "surface elevation changes" is used for the observation, but "thickness change" is use for the model. It seems that only observations on the grounded part are used so that "surface elevation changes" should correspond to "thickness changes", but better discuss this point and check that it is consistent throughout the manuscript.
Thank you for pointing this out. All raised questions have been clarified. A couple of changes/additions to the manuscript:
*"The following seven years are used as calibration period, therefore the temporal mean of the ice thickness change from year four to year ten (inclusive) of the simulations will be compared with satellite observations which also span a seven years period."*
*"Only the last seven years (beginning of 2008 to beginning of 2015) of the dataset are used here for calibration"*
*"There is no exact start date of the simulations which makes a dating of the calibration period difficult. However, the ice flow observations from Rignot at al (2011) used for the ice sheet initialisation are largely from a three year period centered around 2008, which is why this is the first year of surface elevation change observations we use. We do not correct for possible changes in firn thickness and directly convert surface elevation change rates of grounded ice into rates of ice thickness change. An average of all 14 six-month intervals is used for calibration, however for one calibration approach this averaging is performed in basis representation (see Section 3.2 for details)."*

- Finally, I encourage the authors to discuss with more details the benefits of using the surface elevation changes with their experimental design for the calibration. The model is first calibrated using spatial observations of surface velocities to tune the basal friction and viscosity. This point should be made more clear for readers that are not familiar with the initialisation of ice sheet models, i.e. in a sub-section to describe the model initialisation as suggested above. As the ensemble design implies multiplicative perturbations of these inverted fields, the best fit is obtained with the default values (0.5) and all other combinations should degrade the fit to the observed velocities. The fact that the calibration recovers values that are close to the default means that it is the configuration of the model that best fit the velocities that give the best fit to the surface elevation changes. Any other result would mean that the model is not able to reproduce both the velocities and elevation changes, and indicate a problem in the model or in the ensemble design. So a question is how much additional informations do we get from using the surface elevation change field as an additional observation for the model calibration/initialisation? I think it would be difficult to answer this question in a quantitative way, but it would be interesting, in the discussion section, to group and improve the parts discussing the limitation of the experimental design (l 5-10, p18), with the discussion on what we can expect from including the temporal component (l 21-23, p18), even if the calibration period (7 years) seems too short to discriminate the friction law exponent and basal melting scaling.

As suggested we added more information about the inversion in the data description chapter:
*For grounded ice the model inversion attempts to find the optimal combination of the two-dimensional fields of effective viscosity and basal traction coefficients for a given ice geometry to reproduce the before mentioned observed surface speed of the ice. It contains penalty terms to avoid overfitting but does not directly address apparent inconsistencies between the datasets, sometimes framed as 'violations to mass conservation'. In other words, for a given combination of ice geometry and ice speed it is possible that the only way to satisfy mass conservation is by unrealistic, small-scale high-amplitude rates of ice thickness change. These are typically caused by errors in either of the datasets, but interpolation and locally inappropriate model assumptions can contribute as well. The modified bedrock by Nias et al. (2016) is designed to reduce those inconsistencies."*
We agree that there is a certain amount of coupling between surface velocities and thickness change. This coupling is largely due to the ice geometry which makes this approach suitable to address uncertainty in the bedrock. But there are more reasons why the optimized traction and viscosity fields are not perfect (velocity observation errors, limitations of numerical inversions, etc.), which is the motivation to perturb optimized parameters in the first place. Ice thickness change observations can help, as we show, to quantify how much perturbation is necessary to cover the range of reasonable model setups. We tried to make this more clear, including by the following addition:
*"It should also be noted that for a given ice geometry the surface speed (used for initialisation) and ice thickness change (used for calibration) are not fully independent (conversation of mass). Finding the unperturbed traction and viscosity fields to show good agreement with ice thickness change observations is not surprising, yet a good test of the initialisation process, initialisation data and the quality of the initial ice geometry. For the same reasons, the optimized fields cannot be considered without uncertainty. This uncertainty can be quantified by ice thickness change observations, as has been shown here.“*

Minor comments :

- Everywhere ; better to use " friction law " instead of " sliding law ".

done

- Abstract, line 11 : " while a net sea level contribution calibration imposes only weaker constraints ". Maybe not very clear, suggestion " while calibration against an aggregated observation, as the net sea level contribution, imposes only weaker constraints ".

  changed

- Page 2 , line 8 : " basal melting is expected to continue for the next few years to decades ", not sure what do you mean, maybe " high rates of basal melting " ?

  corrected

- Page 5, line 4 : " and use the following 7 years as calibration period " ; see major comment above, explain how the model results are used.

  done

- Page 5, line 4-5 : " Other calibration periods have been tested and show small impact on the results for calibrations in basis representation ", give more details for the meaning of " others " : longer, shorter, different spin-up duration, etc... ?

  done

- Page 5, line 5 : " We regrid the simulated surface elevation fields " ; Please clarify ; is it surface elevation or surface elevation rates of change ?

  The latter, corrected

- Sec. 2.2 Observations : please provide more information on the data that are used. Dates?

  added

- Page 6, line 8-9 : " The first k columns of U) are illustrated in Figure 1 which are related to the PCs (Bi) by multiplication with the singular values. " ; Please check this sentence and how it relates with Eq. (2). It is said in lines 4-5 that the principal components (PCs) are the first columns of B, and caption of Fig.1 says that it shows the 5 PCs.

  We realize that referring to Figure 1 sometimes as normalized PCs and sometimes as U is confusing (even if technically correct since $B_i/\|B_i\| = U$). This has been simplified

- Page 6, line 12-13 : " This decomposition reduces the dimensions from m grid cells to just k principal components. ". B' and V' still have m lines corresponding to the grid cells but k columns, so the dimension reduction is from the n ensemble members to the first k¡¡n PCs ?

  clarified

- Figure 2, caption : " Mean observed ice thickness change ". Date ?

  specified

- Page 7, line 5 : " observations over a seven year period " ; see above ; provide details in Sec. 2.2.

  clarified

- Page 7, line 10-11 : " The spatial variance of the difference between the reprojected and original fields is substantially smaller than from z(xy) alone: ". What are the implications ?

  clarified

- Page 8 : Define that N represent the normal distribution.

  done

- Page 12, line 12 : " fast simulations " ; Needs reformulation : " simulations with high velocities " ?

  done

- Page 12, lines 21-26 : I think this is hardly understandable for a non specialist of ice flow modelling, especially the part " as C compensate for v ". It is maybe better to move Eq. 13 in Sec. 2.1 and give more details there on how the frictions coefficients C are tuned with respect to the observed velocities. Expressed in a simple way, the explanation is that the model has been tuned to give the same initial state, however as the friction laws have a different non-linearity, differences will only become apparent in areas where changes in velocity or stresses are significant. The authors might also be interested by the study from Brondex et al., Sensitivity of centennial mass loss projections of the Amundsen basin to the friction law, Cryosphere, 2019.

  We added more information about the initialization to section 2.1 and tried to make this section easier to understand by removing details (including Eq. 13). It now reads:
  *"The initialization of the ensemble has been performed for each friction law individually which means that the initial speed of the ice is by design equivalent. It is only after the ice velocities change that the different degrees of linearity in the friction law has any impact on the simulations."*

- Page 12, line 29 : " From this test we conclude that basal sliding law and ocean melt scaling cannot be inferred from this calibration approach ". As explained, the problem seems not to be the calibration approach itself but the fact that the changes have not been sufficiently large during the calibration period to distinguish between different sliding laws and different melt scaling.

  We added a note to the calibration period but cannot rule out that other approaches, e.g. one which retains the temporal development, could improve the inference of other parameters

- Page 16, line 15 : " However, no satellite observations have been used for the bedrock modification, nor has there been a quantitative probabilistic assessment. ". Do you mean "radar observations" instead of "satellite observations" ? It would be interesting to compare with the BedMachine bed topography.

clarified

# Point by Point response to review 4

Andreas Wernecke for all authors of the manuscript

This paper presents a series of statistical methods that can be utilised to constrain and supplement simulations of ice flow in order to reduce uncertainty associated with the model initialisation procedure. The authors use statistical emulation to forecast additional simulations within a predetermined parameter space, combined with the use of spatial observations with which to calibrate the emulated ensemble. Thank you for your insightful comments. By addressing these we were able to improve the manuscript in particular regarding the flow of the narrative and interpretation of results. Hearing from the perspective of an ice sheet modeller as potential future user of some of the presented methods is invaluable.

General comments:

- As an ice sheet modeller unfamiliar with some of the statistical methods, I found it challenging to follow the narrative of the various steps involved, what data was being used and why each step was being performed. Each method appears to be presented separately rather than sequentially with continuation from the previous procedure. The sequential process should be more clearly outlined at the beginning of the methods section. A flow diagram would be useful to highlight each procedure and the data used. As a paper presenting a new methodology that could be beneficial to the ice sheet modelling community (by reducing the computational expense required for large ensembles), it is of great importance that the methods are conveyed clearly and can be reproduced by the reader. In its current form this is not the case.

  We followed your advice and added a flow diagram to the manuscript (Figure 1 in this document). The outline of the methods has also been extended with a focus on a clear narrative.

- It is unclear what the purpose of Section 2.2 is. Moreover, the study discusses using 7 years of dh/dt satellite observations, but the dataset is presented as 1992-2015. Which years are used? Why 7?

  Section 2.2 describes the observations used. It has been expanded and clarified in respect to its purpose and how the data is used specifically. Seven years are the time from 2008 (the center of the ice velocity measurements used for initialisation) to the end of the dh/dt dataset (2015). This is now mentioned even though we avoid linking the model period to specific dates to minimize the impression of making predictions.

- On page 12 you propose that the imposed basal melting has a "delayed impact" on the dynamics of the system which is contradictory to a number of studies that highlight ocean forcing to be the primary driver of immediate dynamical response in the ASE. On what timescale do you consider the response to be delayed? Numerous studies have shown that the ASE is highly sensitive to perturbations in ocean driven melting (e.g. Pritchard et al., 2012; Jenkins et al., 2018) whereas you are suggesting that the region is somewhat insensitive and requires considerable melting/thickness change to impact dynamics? This is an important point and it would be worth commenting on this in more detail if such a claim is to be made, or perhaps this should be described more carefully if this is not the case.

We made more clear that the simulations are sensitive to ocean melt and friction law but give the instantaneous impact of other parameters as plausible explanation for dominating the calibration. The suggested references have been incorporated.
*"A change in bedrock, basal traction or viscosity has a much more immediate effect on the ice dynamics. For example, if the basal traction field is halved, the basal drag will be reduced by the same amount leading to a speed up of the ice at the next time step (via the solution of the stress balance). [...] This does not mean that the simulations are insensitive to the ocean melt forcing and friction law, in fact Fig. 4 shows that both parameters have some impact on the simulation in the calibration period. It just means that the much more immediate effects of basal traction and viscosity are likely to dominate the calibration on short time scales."*

- If the weighted average of C and phi are 0.47 and 0.45 this infers that a more slippery bed and softer ice result in better estimates of dh/dt than the optimum (0.5) from the velocity inversion. This should be commented upon. Could this mean that the 0.5 values underestimate sea level contribution?

  Following your comment we added:
  *"While this reduction is relatively small and the central run cannot be ruled out as optimal setup (Likelihood notably larger than zero), this does indicate a possible underestimation of sea level contribution by the default run. With modified bedrock, non-linear friction law and default traction and viscosity values, the SLCs at the end of the simulation period range from 11 to 19.5 mm SLE depending on the ocean melt scaling, while the basis-calibration mean SLC is 19.1 mm SLE (Table 1)."*

- It is unclear what the purpose of the methods performed in section 3.2 are. Why have the observations been reprojected?

  This has been clarified. The purpose of the reprojection is to have the observations and model on the same basis so that they can be compared (by the calibration).

- As was highlighted in the previous round of reviews, I am concerned with the presentation of results as projections of future sea level contribution given the absence of additional forcing throughout the simulation. This should be clarified throughout, and the use of the term "projections" be reconsidered as this is more generally applied to future simulations involving some form of climate forcing. Further, the emphasis of the study should be on the novel methods presented, this, in addition to what the methods employed are, should be more clearly presented in the abstract.

We followed your suggestions and stop using 'projections' to refer to the simulations used here throughout the manuscript. The term "projection" is now only used when it is clear that it is not referring to the simulations used here but instead to highlight the potential value for future applications. The model data chapter now includes the following new paragraph: *"The simulations used here are not intended to be predictions of the future but instead project the current state of the ASE glacial system with a constant recent-past climate forcing and perturbed parameters into the future. No change in the climate is represented in the ensemble. End-of-simulation sea level contribution distributions are presented to illustrate and compare the value of calibrations and should not be understood as best estimates of future sea level contribution."*
We hope this highlights the limitations of the ensemble in very clear terms. We decided to maintain the discussion of climate forcing uncertainty as we feel the need to highlight that there is no robust link between climate change and decadal ocean melt forcing in the ASE. Therefore we think that the constant forcing is no compelling reason to discard these simulations as toy example/unrealistic.

- In the previous round of reviews it was suggested that the emphasis of the paper should be on the use of new statistical methods for model calibration and emulation, instead of the 50 year projections of SLE that arise from the investigation. Whilst you begin to do this by emphasising in the discussion that future ocean forcing of the ASE will accelerate the dynamic response of the region, you then contradict this by stating "climate scenarios would have a small net impact on our 50-year projections".

  The reason why a potentially increased dynamic mass-loss is not a contradiction to a small net mass-loss is the surface mass balance which is likely to have a more negative impact on global mean sea level for high emission scenarios. Warmer air masses are able to transport more humidity which is expected to increase the surface accumulation in Antarctica. To be clear, we do only state that "future ocean forcing of the ASE will accelerate the dynamic response of the region" in the sense that if the ocean temperature happens to increase this would entail an accelerated dynamic response. We do not feel confident to make any predictions about decadal ocean melt forcing for the ASE.

  Studies have indicated that the range of possible forcings within the RCP scenarios could have a substantial impact on the response of the region over a 50 year period as simulations have shown that the region responds linearly to ocean melting (see Alevropoulos-Borrill et al. 2019).
  Agreed, but we note that this substantial impact is nearly entirely due to variability and not long term trends. The perturbation of melt forcing in the ensemble used here also has a substantial impact on the simulations.
  Furthermore, Alevropoulos-Borrill et al. (2019) find that region becomes more sensitive to the perturbed model parameters (investigated by Nias et al. 2016) as the ocean forcing increases and therefore climate scenarios would impact the projections in this investigation. Future studies would benefit from the application of the method presented in this investigation to climate forced projections and this should be highlighted in the discussion.

We made this more clear in the discussion, including:
*"The method presented here can be applied to forced simulations which would benefit from reduced uncertainty intervals to highlight the impact of climate change on ice sheet models."*
and
*"Similar improvements should be achievable for ice sheet simulations forced by global climate model projections."*

- Continuing from the previous point, mentioning the large uncertainties associated with future ocean forcing and the implementation of basal melting in ice sheet models is a viable point however the relevance of this to why you apply no additional forcing to your simulations is unclear. If the uncertainty associated with ocean forcing is so wide, is this really captured in a halving and doubling of the optimal ocean melting obtained during the initialisation procedure?

  Not being able to define robust representative forcings can justify to use a simplified forcing instead. This point is supported by Alevropoulos-Borrill et al. (2019) where two of the RCP 8.5 climate models imply forcings which leads to SLCs lower than the control run through most of the model period (Figure 4). Climate models have such a strong impact and there is (to our knowlage) no commonly agreed upon way to select which one to use. Since this study does not focus on this topic, it seems hard to justify to choose any specific climate model for the forcing (making the impression to know the future forcing while we do not), and hence choosing to use none of them is a legitimate option. That discussion is intended to highlight that the simulations used here are not to be discarded as unrealistic/a toy example but being somewhere in the realm of forced projections. This is highlighted to show that the results from this study are likely to be transferable to forced projections. We do not know the appropriate range of perturbations to the melt forcing. Unfortunately our method has limited success in constraining it (which would as well be the case if the perturbations used here are too small, but we do not feel confident to imply this in the manuscript). Varying it by a factor of four is more than some other studies do.

- I believe the figures were modified following the previous round of reviews but these updated figures were not included in the revised manuscript. This should be amended and avoided in the future.

  We are sorry to hear that. The figures have indeed been updated in the last revision and have been included in the file we uploaded. We did double-check the correct appearance in the current upload.

Specific comments:

- Page 1: Given that the investigation is heavily methods focused, the abstract does not fully convey the methods employed which should be more clearly stated for the reader (this is a more important focus than the 50 year SLE "projections").

  adjusted

- Page 1 line 1: Calibration of what with observations?

  model simulations, clarified

- Page 1 line 2: "...particularly if this exploits as much of the available information as possible (such as spatial characteristics)" seems vague?

  More details are given in the following lines. This line is intended to introduce the larger setting of the challenges addressed in this study

- Page 2: The introduction, particularly the first paragraph is very long-could this be shortened and maintain a more study-relevant focus?

  done

- Page 2 line 12: Consider removing the clause "centered at the Ellsworth Mountains".

  done

- Page 2 line 23: Unclear why the sentences in brackets are relevant.

  To define the meaning of 'input parameters' for this study, clarified

- Page 2 line 29: This sentence could be more concise.

  done

- Page 2 line 31: Remove ie and replace with "in order" or equivalent.

  done

- Page 2 line 31: Why does reducing uncertainties matter? This should be included in the introduction.

  We refer to the fact that current projections are only just so distinguishable from zero and infer from it the need for more precise projections. We tried to make this point more clear by adding:
  *"In other words, the uncertainties are of the same order of magnitude as the projections themselves, hence the reduction of uncertainty is essential to quantify projections effectively."*
  We believe that readers of 'The Cryosphere' are familiar with the need to make precise projections of future sea level rise, without e.g. repeating the number of people living close to coastlines.

- Page 3 line 10: Rethink paragraphing of this as the beginning sentence better fits with the previous paragraph.

  done

- Page 3 line 15: "in the following section"

  Sorry, we do not understand what you mean. Adding "in the following section" to this line seems inappropriate.

- Page 4 line 25: "...Hypercube design by Nias et al., (2016)."

  corrected

- Page 5 line 3: Move "For a full description of the model..." to a different paragraph

  done

- Page 5 line 10: “We use a compilation of five satellite altimeter datasets of surface elevation changes...” to do what?

  for calibration, added

- Page 5 line 21: “to represent”

  corrected

- Page 5 line 31: Principal Component Decomposition in section 3.1 is performed on what, the whole Nias ensemble?

  yes, clarified

- Page 6 figure 1: The figure caption is vague. The modes of variation of what variable? Within the Nias ensemble? Relative to the dh/dt observations? No scale bar label.

  clarified

- Page 6 line 8: U)? Typo?

  yes, corrected

- Page 6 line 9: Sometimes you have Figure 1 sometimes Fig. 1 in the text. Make these consistent.

  corrected

- Page 7 figure 2: Mean observed ice thickness change using which dataset? Over what time period? In what way have the observations been reprojected, there is little discussion of this in the text and it is unclear what the purpose of this figure is to the reader. The caption should be less vague.

  clarified

- Page 7 line 6. Does this sentence mean the observations are assumed to be temporally constant over the 7 year period?

  It means that the temporal development of the observations is not captured. The differences in observation within the seven years do emerge in the observational uncertainty

- Page 7 line 9: five in letters

  corrected

- Page 8 line 25: lowercase S

  corrected

- Page 9 line 28: “can” in the wrong place

  corrected

- Page 9 line 29: rephrase sentence and remove e.g.

  done

- Page 10 line 25: If 1.4% of the parameter space cannot be ruled out, does this mean you discard 98.6% of the emulated parameter sets?

  Yes. Those 98.6% have virtually no likelihood attributed to them so that the history matching has no impact on the likelihood distributions. Above everything, this is a test whether all setups are ruled out or not.

- Page 10 line 6-7: rethink paragraphing

  After careful consideration we decided to leave this unchanged

- Page 10 line 29: Both using Eq. and Equation in the text. Choose one and make it consistent.

  corrected

- Page 11 line 9: "These 14 realizations are used in exactly the same way as described before..." this could be more clear.

  changed

- Page 11 line 13: "many other" could you not specify how many additional tests you explored?

  good point, done (11 are shown in the supplement)

- Page 12 line 2: What does "good model configuration" mean?

  rephrased

- Page 12 line 11: If the linear sliding law is favoured due to the density of central ensemble members, shouldn't this be presented as a caveat of the method? Are there any methods that would help to identify such a biasing of results?

  We consider this a caveat of the ensemble and not so much of the calibration. All calibration approaches tested here have the same problem in identifying the correct sliding law. Identifying which parameters are well constrained by synthetic model tests is probably the best method to identify such issues. In cases like this it comes down to the parameter prior to represent the experts believe. We added: *"This can be considered a caveat of the model ensemble which might very well be present in other ensembles which perturb the friction law in combination with other parameters. If the friction law cannot be adequately constrained, as is the case for all calibration approaches tested here, the prior believe in the optimal friction law must be set very carefully."*

- Page 12 line 12: Clarify what you mean by 'fast' and 'slow' simulations

  done

- Page 12 line 16: Your editor response gives 32% to 68% whilst the revised manuscript gives 28% to 72%.

  Within the last round of reviews several changes to the analysis have been proposed. While we used the original analysis as default case in the response to the editor, the fully updated analysis is used for the manuscript. If we remember correctly it is the use of a spin-up period which had the largest impact on these numbers. We can confirm that 28% to 72% is correct for the latest version.

- Page 12 line 17: Rhetorical question unnecessary, reword.Page 12 line 30: In Nias (2017; Ph.D. thesis) it is suggested that the halving and doubling of the initial melt rates did not capture a wide enough range. It might be worth mentioning this. Page 13 line 23: Reconsider paragraphing of this.

  Question has been rephrased and paragraphing changed

- Page 13 line 25: Should this be viscosity parameter not velocity?

  yes, corrected

- Page 14 line 1: Grey and Brown need not be in capitals.

  corrected

- Page 14 line 3: wile? Typo?

  corrected

- Page 15 line 1: 6mm SLC

  corrected

- Page 15 line 1: Inconsistency with SLE and SLC, choose one and stick with it throughout.

  We use Sea level equivalent (SLE, typically mm SLE) as unit and sea level contribution as quantity name (example: The SLC is 6 mm SLE)

- Page 15 figure 5b: The grey shading makes it difficult to read the histogram.

  The grey shading is the emulator histogram. After careful consideration we decided to leave this unchanged.

- Page 17 line 9: Two commas.

  corrected

- Page 17 line 25: See also Alevropoulos-Borrill et al. (2019).

  Alevropoulos-Borrill et al. (2019) is supporting this statement but we can only reference accepted peer reviewed papers

- Page 17 line 23: "Figure 2"

  corrected

- Page 18 line 6: What sort of variations are performed in the cited papers?

  added

- Page 18 line 8: "Probabilistic calibrations are an assessment of model setups to be the best of all tested cases" this sentence is unclear.

  rephrased

- Page 18 line 11: Consider moving paragraph to conclusions or beginning of discussion.

  done

- Page 18 line 31: As mentioned in the general comments, claiming that ocean melting has a slow impact on ice sheet behaviour is ambiguous and the author should be more careful with wording such a statement.

  agreed, this statement has been reworded in the discussions and removed from the conclusion

- Page 18 line 26: Given that the simulations include no future climate forcing, is it realistic to present the findings as the next 50 years. The absence of climate forcing should be more clearly highlighted here.

  done

- Page 19 line 6: Final paragraph in the conclusion conveys that the estimates are "projections" and does not include the fact that there is no additional forcing applied in these simulations. This is misleading.

  corrected

References: Alevropoulos-Borrill, A. V., Nias, I. J., Payne, A. J., Golledge, N. R., and Bingham, R. J.: Ocean forced evolution of the Amundsen Sea catchment, West Antarctica, by 2100, The Cryosphere Discuss., https://doi.org/10.5194/tc-2019-202, in review, 2019.

Pritchard, H., Ligtenberg, S.R., Fricker, H.A., Vaughan, D.G., van den Broeke, M.R. and Padman, L., 2012. Antarctic ice-sheet loss driven by basal melting of ice shelves. Nature, 484(7395), pp.502-505.

added

Jenkins, A., Shoosmith, D., Dutrieux, P., Jacobs, S., Kim, T.W., Lee, S.H., Ha, H.K. and Stammerjohn, S., 2018. West Antarctic Ice Sheet retreat in the Amundsen Sea driven by decadal oceanic variability. Nature Geoscience, 11(10), pp.733-738.

added

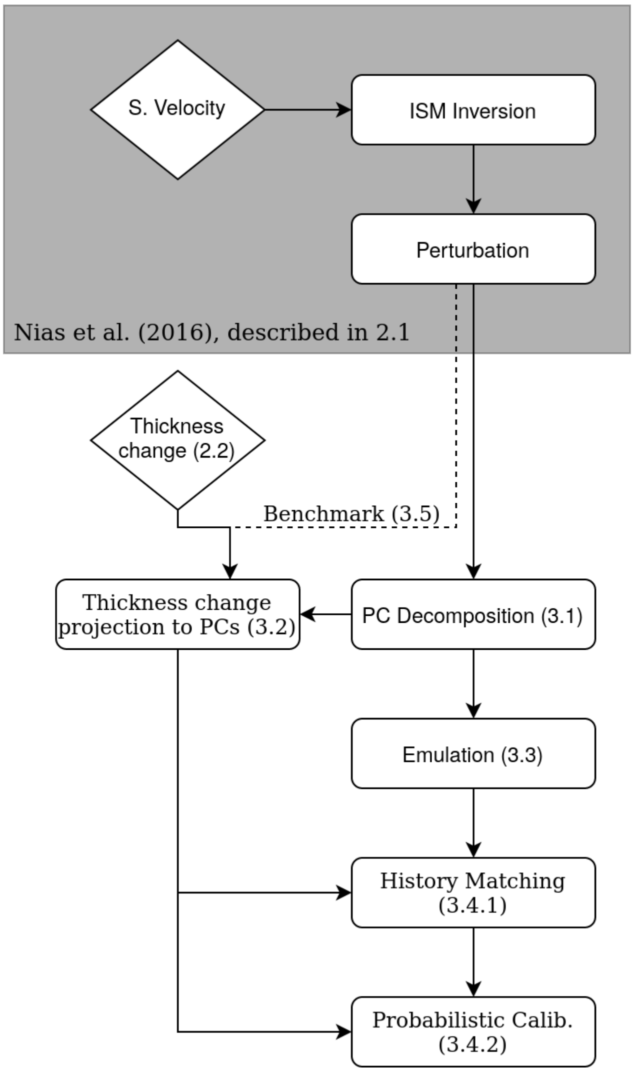


Figure 1: Flow diagram of the proposed calibration procedure. Horizontal boxes represent steps in the analysis, diamonds represent observations and numbers refer to corresponding Sections in this study

# Spatial probabilistic calibration of a high-resolution Amundsen Sea Embayment ice-sheet model with satellite altimeter data

Andreas Wernecke[1], Tamsin L. Edwards[2], Isabel J. Nias[3,4], Philip B. Holden[1], and Neil R. Edwards[1]

[1]School of Environment, Earth and Ecosystem Sciences, The Open University, Milton Keynes, UK
[2]Department of Geography, King's College London, London, UK
[3]Earth System Sciences Interdisciplinary Center, University of Maryland, College Park, MD, USA
[4]Cryospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA

**Correspondence:** Andreas Wernecke (andreas.wernecke@open.ac.uk)

**Abstract.**

Probabilistic predictions of the sea level contribution from Antarctica often have large uncertainty intervals. Calibration of model simulations with observations can reduce uncertainties and improve confidence in projections, particularly if this exploits as much of the available information as possible (such as spatial characteristics), but the necessary statistical treatment is often challenging and can be computationally prohibitive. Ice sheet models with sufficient spatial resolution to resolve grounding line evolution are also computationally expensive.

Here we address these challenges by adopting ~~a novel~~ and comparing dimension-reduced ~~approach to calibration , combined with statistical emulation~~ calibration approaches based on a principal component decomposition of the adaptive mesh model BISICLES. The effects model parameters have on these principal components are then gathered in statistical emulators to allow for smooth probability density estimates. With the help of a published perturbed parameter ice sheet model ensemble of the Amundsen Sea Embayment (ASE), we show how the use of principal components in combination with spatially resolved observations can improve probabilistic calibrations. In synthetic model experiments (calibrating the model with altered model results) we can identify the correct basal traction and ice viscosity scaling parameters as well as the bedrock map with spatial calibrations~~while a~~. In comparison a simpler calibration against an aggregated observation, the net sea level contribution~~calibration ,~~, imposes only weaker constraints by allowing a wide range of basal traction and viscosity scaling factors.

~~Projections of the 50 year~~ Uncertainties in sea level rise contribution of 50 year simulations from the current state of the ASE can be ~~narrowed down~~ reduced with satellite observations of recent ice thickness change ~~to~~ by nearly 90%; Median and 90% confidence intervals are 18.9 [13.9, 24.8]~~mm (median and 90% range) with~~ mm SLE for the proposed spatial calibration approach, ~~compared to~~ 16.8 [7.7, 25.6] mm SLE for the net sea level calibration and 23.1 [-8.4, 94.5] mm SLE for the uncalibrated ensemble. The spatial model behaviour is much more consistent with observations if, instead of Bedmap2, a modified bedrock topography is used that most notably removes a topographic rise near the initial grounding line of Pine Island Glacier.

The ASE dominates the current Antarctic sea level contribution, but other regions have the potential to become more important on centennial scales. These larger spatial and temporal scales would benefit even more from methods of fast but exhaustive model calibration. ~~Our approach therefore has the potential to improve projections for the~~ Applied to projections of the whole

Antarctic ice sheet~~on continental and centennial scales by efficiently improving~~ , our approach has therefore the potential to efficiently improve our understanding of model behaviour, ~~and~~ as well as substantiating and reducing projection uncertainties.



## 1 Introduction

The Antarctic ice sheet is currently losing mass at a rate of around 0.5 to 0.6 mm ~~/yearsea level equivalent~~global mean Sea Level Equivalent (SLE) per year, predominantly in the Amundsen Sea Embayment (ASE) area of the West Antarctic Ice Sheet (WAIS) (Shepherd et al., 2018; Bamber et al., 2018). This is due to the presence of warm Circumpolar Deep Water causing sub-shelf melting and ice dynamical changes including retreat of the grounding line that divides grounded from floating ice (Khazendar et al., 2016). ~~The dynamical changes are consistent with those expected from the Marine Ice Sheet Instability (MISI) hypothesis (Favier et al., 2014; Ritz et al., 2015). Although projections of future ocean changes are uncertain, basal melting is expected to continue for the next few years to decades, possibly even if the external oceanic heat flux towards the ice sheet decays (Naughten et al., 2018). Persistent grounding line retreat could lead eventually to a collapse of the marine-based WAIS, contributing up to 3.4 m equivalent to global mean sea level (Fretwell et al., 2013) even though there are indications that a small part of the WAIS, centered at the Ellsworth Mountains, existed at least for the last 1.4 million years (Hein et al., 2016).~~ However, the future response of the Antarctic ice sheet to a changing climate is one of the least well understood aspects of climate predictions (Church et al., 2013). Predictions of the dynamic ice sheet response are challenging because local physical properties of the ice and the bedrock it is laying on are poorly observed. Parameterisations of unresolved physical processes are often used and need to be validated (DeConto and Pollard, 2016; Edwards et al., 2019; Cornford et al., 2015; Pattyn et al., 2017). Progress has been made in the ~~understanding of ice sheet feedbacks, like MISI and the Marine Ice Cliff Instability hypothesis (DeConto and Pollard, 2016), as well as the~~ development of numerical models with higher resolutions and improved initialization methods (Pattyn, 2018). But these improvements cannot yet overcome the challenges of simulating what can be described as under-determined system with more unknowns than knowns. For this reason, some studies use parameter perturbation approaches which employ ensembles of model runs, where each ensemble member is a possible representation of the ice sheet using a different set of uncertain input parameter values (Nias et al., 2016; DeConto and Pollard, 2016; Schlegel et al., 2018; Gladstone et al., 2012; Ritz et al., 2015; Bulthuis et al., 2019) (~~Here we do not distinguish between~~ In this context 'input parameters' can refer to initial values of state variables, which will change during the simulation, ~~and~~ or model parameters, which represent physical relationships. All of those quantities can be poorly known and contribute to uncertainties in predictions.). In most studies, the computational expense of exploring uncertainties either restricts the minimum spatial resolution to several kilometres, causing challenges in representing the grounding line, or else ~~are restricted regional applications~~restricts the application to regional scale. One exception is the ensemble by Nias et al. (2016), which uses the adaptive mesh model BISICLES at sub-km minimum resolution over the ASE domain (Pine Island, Thwaites, Smith and Pope glaciers).

In Antarctic ice sheet model ensemble studies, the projected sea level contribution for high emission scenarios by the end of the century typically ranges from ~~around zero to tens of~~ about zero to about 40 centimetres, i.e. the ensemble spread (~40 cm) is twice the predicted (mean/median) contribution (~20 cm) (Edwards et al., 2019). It is therefore essential to constrain ice sheet model parameters to reduce these uncertainties ~~i.e.~~ in order to attain sharper and more distinctive prediction distributions for different climate scenarios. In other words, the uncertainties are of the same order of magnitude as the projections themselves, hence the reduction of uncertainty is essential to quantify projections effectively. Statistical calibration of model parameters refines predictions by using observations to judge the quality of ensemble members, in order to increase confidence in, and potentially reduce uncertainty in, the predicted distributions. Calibration approaches range from straightforward ‘tuning’ to formal probabilistic inference. Simple ruled out/not ruled out classifications (also called history matching or precalibration) can be used to identify and reject completely unrealistic ensemble members while avoiding assumptions about the weighting function used for the calibration (e.g. Holden et al., 2010; Williamson et al., 2017; Vernon et al., 2010). Formal probabilistic, or Bayesian, calibrations using high dimensional datasets require experience of statistical methods and can be computationally prohibitive (Chang et al., 2014). There are few ice sheet model studies using calibrations, among which are history matching (DeConto and Pollard, 2016; Edwards et al., 2019), gradual weight assignments (Pollard et al., 2016) and more formal probabilistic treatments (Ritz et al., 2015; Chang et al., 2016b, a). Most use one or a small number of aggregated summaries of the observations, such as spatial and/or temporal averages, thus discarding information that might better constrain the parameters.

Ideally, then, calibrating a computer model with observations should use all available information, rather than aggregating the observations with spatio-temporal means.

However, the formal comparison of model simulations with two-dimensional observations, such as satellite measurements of Antarctica, poses statistical challenges. Measurements of the earth system typically show coherent spatial patterns, meaning that nearby observations are highly correlated due to the continuity of physical quantities. Model to observation comparisons on a grid-cell-by-grid-cell basis can therefore not be treated as statistically independent. On the other hand, appropriate treatment of these correlations with the inclusion of a co-variance matrix in the statistical framework for calibration can be computationally prohibitive (Chang et al., 2014). While the simplest way to avoid this is by aggregation, either over the whole domain (Ritz et al., 2015; DeConto and Pollard, 2016; Edwards et al., 2019) or subsections assumed to be independent (Nias et al., 2019), a more sophisticated approach that preserves far more information is to decompose the spatial fields into orthogonal Principal Components (PCs) (Chang et al., 2016a, b; Holden et al., 2015; Sexton et al., 2012; Salter et al., 2018; Higdon et al., 2008). The decompositions are used as simplified representations of the original model ensemble in order to aid predicting the behaviour of computationally expensive models, and in some cases to restrict flexibility of the statistical model in parameter calibration so that the problem is computationally feasible and well-posed (Chang et al., 2016a, b). But the latter studies, which employ a formal probabilistic approach, still assume ~~spacial~~ spatial and/or temporal independence at some point in the calibration. This independence assumption is not necessary if the weighting (likelihood) calculation is shifted from the spatio-temporal domain into that of principal component basis vectors, as proposed e.g. in Chang et al. (2014).

A further difficulty is the computational expense of Antarctic ice sheet models that have sufficient spatial resolution to resolve grounding line migration. This can be overcome by building an 'emulator', which is a statistical model of the response of a physically-based computer model. Emulation allows a small ensemble of the original ice sheet model to be extended to a much larger number. This approach has recently been applied in projections of the Antarctic ice sheet contribution to sea level rise by interpolation in the input parameter space in general (Edwards et al., 2019; Chang et al., 2016a, b; Bulthuis et al., 2019) and melt forcing in particular (Levermann et al., 2014). Emulation becomes particularly important in model calibration, as this down-weights or rejects ensemble members and therefore reduces the effective ensemble size.

The aim of this study is to ~~use a novel,~~ develop a practical, yet comprehensive calibration ~~of~~ approach for data from the high-resolution ~~Antarctic~~ ice sheet model BISICLES~~to give smooth, refined probability~~. This approach is compared to more traditional methods by means of a synthetic model test and the impact on probability density functions for the dynamic sea level contribution from 50 year simulations of the Amundsen Sea Embayment~~for 50 years from the present day~~. We derive principal components of ice thickness change estimates with a singular value decomposition, thus exploiting more of the available information of satellite observations than previous studies. The statistical independence of those PCs aids the ~~of~~ use of Bayesian (probabilistic) inference. We use emulation of the ice sheet model to ensure dense sampling of the input space and therefore smooth probability density functions. ~~Emulating the full spatial fields allows us to assess the probabilities not only of total mass loss (in mm Sea Level Equivalent, SLE) but also of the locations of grounding line retreat.~~

In Section 2 we describe the ice sheet model and satellite observation data, followed by ~~our calibration approach~~ the introduction of the calibration approaches used and the benchmark procedure in Section 3. In Section 4 we present the ~~resulting~~ benchmark results and probabilistic ice sheet ~~projections~~ simulation distributions which are discussed in Section 5.

## 2 Model Ensemble and Observations

### 2.1 Ice sheet model ensemble

#### 2.1.1 Ensemble setup

We use the ice sheet model ensemble published in Nias et al. (2016) using the adaptive mesh model BISICLES (Cornford et al., 2013) with equations from Schoof and Hindmarsh (2010). The mesh has a minimum spatial resolution of 0.25 km and evolves during the simulation. The model was run for the Amundsen Sea Embayment with constant climate forcing for 50 years with 284 different parameter configurations. Two uncertain inputs are varied categorically: two different bedrock elevation maps are used, as well as two different ~~sliding~~ friction law exponents. The first bedrock elevation map is Bedmap2, which is based on an extensive compilation of observations (Fretwell et al., 2013), while the second was modified by Nias et al. (2016) in order to reduce unrealistic model behaviour. The modifications are primarily local (<10 km) and include the removal of a topographic rise near the initial grounding line of Pine Island Glacier. The ~~sliding~~ friction law exponent defines the linearity of the basal ice velocity with basal traction, and values of 1 (linear) and 1/3 (power law) have been used. In addition, three scalar parameters were perturbed continuously, representing amplitude scalings of (1) the ocean-induced basal melting underneath ice shelves (i.e. the floating extensions of the ice streams), (2) the effective viscosity of the ice, determining the dynamic response to horizontal strain, and (3) the basal traction coefficient representing bedrock-ice interactions and local hydrology. The default values for these three parameters were determined for initialisation by model inversion (Habermann et al., 2012; MacAyeal et al., 1995) of surface ice speeds ~~(Rignot et al., 2011b), and~~ from Rignot et al. (2011a). For grounded ice the model inversion attempts to find the optimal combination of the two-dimensional fields of effective viscosity and basal traction coefficients for a given ice geometry to reproduce the before mentioned observed surface speed of the ice. It contains penalty terms to avoid over-fitting but does not directly address apparent inconsistencies between the datasets, sometimes framed as 'violations to mass conservation'. In other words, for a given combination of ice geometry and ice speed it is possible that the only way to satisfy mass conservation is by unrealistic, small-scale high-amplitude rates of ice thickness change. These are typically caused by errors in either of the datasets, but interpolation and locally inappropriate model assumptions can contribute as well. The modified bedrock by Nias et al. (2016) is designed to reduce those inconsistencies.

The scaling parameters are subsequently perturbed between half and double the default values in a Latin Hypercube design by ~~(Nias et al., 2016)~~Nias et al. (2016). Different default basal traction coefficient fields have been found for each combination of bed topography and ~~sliding~~ friction law while the default viscosity field only differs between bed geometries (but not ~~sliding~~ friction laws). We use the normalized parameter ranges with halved, default and doubled scaling factors mapped to 0, 0.5 and 1, respectively.

### 2.1.2 Ensemble behaviour

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~~/~~ SLE per year and 1.62 mm~~/~~ SLE per year, respectively. About 10% of the ensemble shows an increasing volume above flotation (negative sea level contribution) and the central runs (0.5 for traction, viscosity and ocean melt parameters) contribute 0.27~~mm/~~ mm SLE per year (linear ~~sliding~~friction) and 0.26~~mm/~~ mm SLE per year (nonlinear ~~sliding~~friction). The average contributions are generally reasonably close to satellite observations (~~0.33 $\pm$ 0.05 mm/~~0.33 $\pm$0.05 mm SLE per year from 2010-2013 (McMillan et al., 2014)) with 0.30~~mm/~~ mm SLE per year for linear ~~sliding~~ friction and modified bedrock, 0.37 mm~~/~~ SLE per year for linear ~~sliding~~ friction and Bedmap-2, 0.38 mm~~/~~ SLE per year for nonlinear ~~sliding~~ friction and modified bedrock and 0.51 mm~~/~~ SLE per year for nonlinear ~~sliding~~ friction and Bedmap-2 (Nias et al., 2016).

~~For a full description of the model ensemble see Nias et al. (2016).~~ We allow for a short spin up phase of 3 years (selected by manual inspection) for the model to adjust to the perturbations~~and use the following 7 years~~. The following seven years are used as calibration period~~. Other~~, therefore the temporal mean of the ice thickness change from year four to year ten (inclusive) of the simulations will be compared with satellite observations which also span a seven years period.

Other spin-up and calibration periods have been tested and show small impact on the results for calibrations in basis representation. For example the median for the basis-calibration of the sea level contribution at the end of the simulations is 18.9 mm SLE with the described three year spin-up and seven year calibration period and 19.1 mm SLE for a seven year spin-up followed by a short three year calibration period. We further tested three year spin-up with four year calibration period and other calibration approaches (see supplement).

We regrid the simulated ~~surface elevation~~ ice thickness change fields for this period to the same spatial resolution as the observations (10 km×10 km) by averaging. ~~The~~ Estimates of the sea level rise contribution at the end of the model period (50 years), used to illustrate the impact of calibrations on simulations of the future, is calculated directly on the model grid~~, using~~ . We use the same catchment area mask as in Nias et al. (2016).

The simulations used here are not intended to be predictions of the future but instead project the current state of the ASE glacial system with a constant recent-past climate forcing and perturbed parameters into the future. No changes in the climate are represented in the ensemble. End-of-simulation sea level contribution distributions are presented to illustrate and compare the value of calibrations and should not be understood as best estimates of future sea level contribution. For a full description of the model ensemble see Nias et al. (2016).

### 2.2 Observations

~~We use~~ The calibration target is based on a compilation of five satellite altimeter datasets of surface elevation changes from 1992-2015 by Konrad et al. (2017). The synthesis involves fitting local empirical models over spatial and temporal extents of up to 10 km and 5 years, respectively, as developed by McMillan et al. (2014). The satellite missions show high agreement, with a median mis-match of 0.09 m/year. The dataset has a resolution of 10 km×10 km spatially and six month temporally. Only the last seven years (beginning of 2008 to beginning of 2015) of the dataset are used here for calibration. The following satellite missions contributed to this period: ERS-2 (until 2011), Envisat (until 2012), ICESat (until 2009) and CryoSat-2 (2010 to 2015). All of these carry radar altimeters, the only exception being ICESat, which had a Laser Altimeter (lidar) as payload.

There is no exact start date of the simulations which makes a dating of the calibration period difficult. However, the ice flow observations from Rignot et al. (2011a) used for the ice sheet initialisation are largely from a three year period centered around 2008, which is why this is the first year of surface elevation change observations we use. We do not correct for possible changes in firn thickness and directly convert surface elevation change rates of grounded ice into rates of ice thickness change. An average of all 14 six-month intervals is used for calibration, however for one calibration approach the averaging is performed in basis representation (see Section 3.2 for details).

## 3 Theoretical basis and Calibration Model

In the following we propose a new ice sheet model calibration approach~~which~~ , as outlined in Fig. 1. It will be tested in section 3.5 and compared to alternative approaches in section 3.6. This calibration approach consists of an ~~emulation and a calibration step. Emulation - statistical modelling of the ice sheet model - helps to overcome computational constraints and to refine probability density functions, while the subsequent calibration infers model parameter values which are likely to lead to good representations of the ice sheet. Both~~ initial spatial decomposition of the model data into Principal Components (PCs) which strongly simplifies subsequent emulation and calibration~~take place in the basis representation of a Principal Component (PC) decomposition, in order~~ . In particular it helps to adequately represent spatial correlation and avoid unnecessary loss of information (e.g. by comparing total or mean model-observation differences). Emulation - statistical modelling of the ice

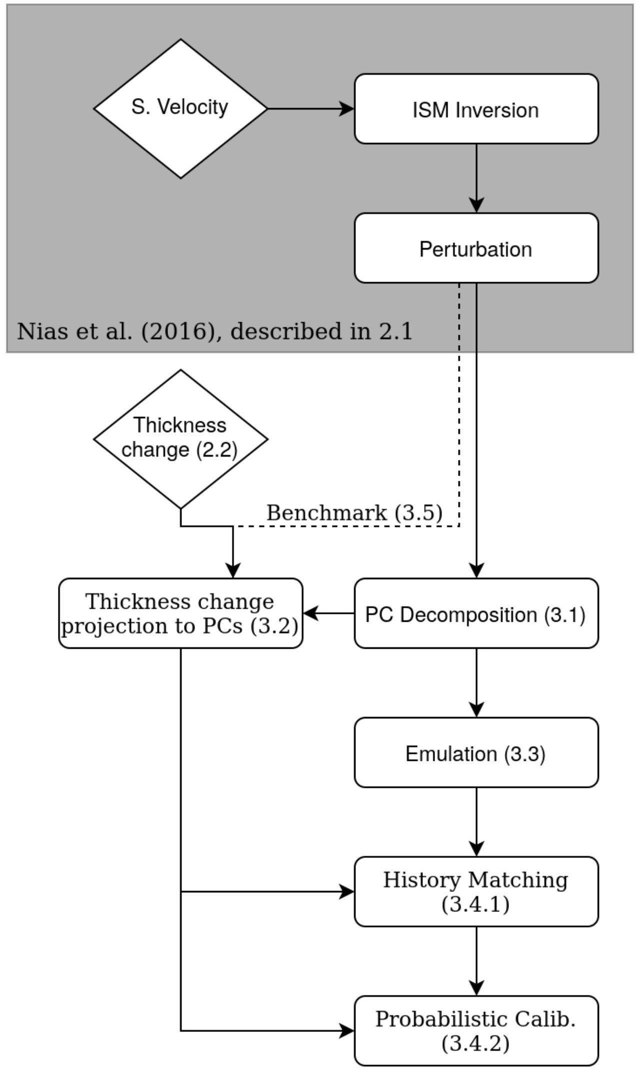


**Figure 1.** Flow diagram of the proposed calibration procedure. Horizontal boxes represent steps in the analysis, diamonds represent observations and numbers refer to corresponding Sections in this study

sheet model - helps to overcome computational constraints and to refine probability density functions. We construct a spatial emulator for ice thickness change in the calibration period to ~~represents~~ represent the two dimensional model response~~in ice thickness change. A~~. In this way we predict how BISICLES would behave for additional perturbed-parameter runs, and use the much larger emulator ensemble in the subsequent calibration instead of the original BISICLES ensemble. The calibration then infers model parameter values which are likely to lead to good representations of the ice sheet. These parameter probabilities are used as weights for a second, non-spatial emulator ~~represents~~ to represent the total sea level rise at the end of the 50 year simulations.

### 3.1 Principal Component Decomposition

Let $\boldsymbol{y}(\boldsymbol{\theta_i})$ be the $m$ dimensional spatial model ice thickness change output for a parameter setting $\boldsymbol{\theta_i}$, where $m$ is the number of horizontal grid cells and the model ensemble has $n$ members so that $\boldsymbol{\theta_1,...,\theta_n} = \boldsymbol{\Theta}$, $\boldsymbol{\Theta} \subset [0,1]^d \subset R^d$ being the whole set of input parameters, spanning in our case the $d = 5$ dimensional model input space. The $m \times n$ matrix $\widetilde{\mathbf{Y}}$ is the row-centered combined model output of the whole Nias et al. (2016) ensemble with the $i$.th column consisting of $\boldsymbol{y}(\boldsymbol{\theta_i})$ minus the mean of all ensemble members, $\bar{\boldsymbol{y}}$, and each row represents a single location. In the following we will assume $n < m$. A principal component decomposition is achieved by finding $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$ so that

$$\widetilde{\mathbf{Y}} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{1}$$

where the $m \times n$ rectangular diagonal matrix $\mathbf{S}$ contains the $n$ positive singular values of $\widetilde{\mathbf{Y}}$ and $\mathbf{U}$ and $\mathbf{V}^T$ are unitary. The rows of $\mathbf{V}^T$ are the orthonormal eigenvectors of $\widetilde{\mathbf{Y}}^T\widetilde{\mathbf{Y}}$ and the columns of $\mathbf{U}$ are the orthonormal eigenvectors of $\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^T$. In both cases the corresponding eigenvalues are given by $diag(\mathbf{S})^2$. By convention $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}^T$ are arranged so that the values of $diag(\mathbf{S})$ are descending. We use $\mathbf{B} = \mathbf{U}\mathbf{S}$ as shorthand for the new basis and call the $i$.th column of $\mathbf{B}$ the $i$.th principal component. The first five principal components have been normalized $(\frac{\mathbf{B}_i}{|\mathbf{B}_i|})$ for Fig. 2 to show more detail of the spatial pattern.

The fraction of ensemble variance represented by a principal component is proportional to the corresponding eigenvalue of $\mathbf{U}$ and typically there is a number $k < n$ for which the first $k$ principal components represent the whole ensemble sufficiently well. We choose $k = 5$ so that 90% of the model variance is captured (~~Figure 2). The first k columns of U) are illustrated in Figure 2which are related to the PCs ($\mathbf{B}_i$)by multiplication with the singular values~~Fig. 2).

$$\widetilde{\mathbf{Y}} \approx \mathbf{B}'\mathbf{V}'^T \tag{2}$$

with $\mathbf{B}'$ and $\mathbf{V}'$ consisting of the first $k$ columns of $\mathbf{B}$ and $\mathbf{V}$.

This ~~decomposition reduces the dimensions from $m$ grid cells to just $k$ principal components~~truncation limits the rank of $\widetilde{Y}$ to $k = 5$. The PCs are by construction orthogonal to each other and can be treated as statistically independent.

### 3.2 Observations in basis representation

One of the calibration approaches we investigate uses the PCs derived before for both the model and observations (see Section 3.4). For this we have to put the observations onto the same basis vectors (PCs) as the model data. Spatial $m$ dimensional observations $\boldsymbol{z_{(xy)}}$ can be transformed to the basis representation by:

$$\hat{\boldsymbol{z}} = (\mathbf{B}'^T\mathbf{B}')^{-1}\mathbf{B}'^T\boldsymbol{z_{(xy)}} \tag{3}$$

for $\boldsymbol{z_{(xy)}}$ on the same spatial grid as the model output $\boldsymbol{y}(\boldsymbol{\theta})$ which has the mean model output $\bar{\boldsymbol{y}}$ subtracted for consistency.

We perform the transformation as in ~~Equation 3~~Eq. (3) for all of the bi-yearly observations over a seven year period to get 14 different realizations of $\hat{\boldsymbol{z}}$. Due to the smooth temporal behaviour of the ice sheet on these timescales we use the observations

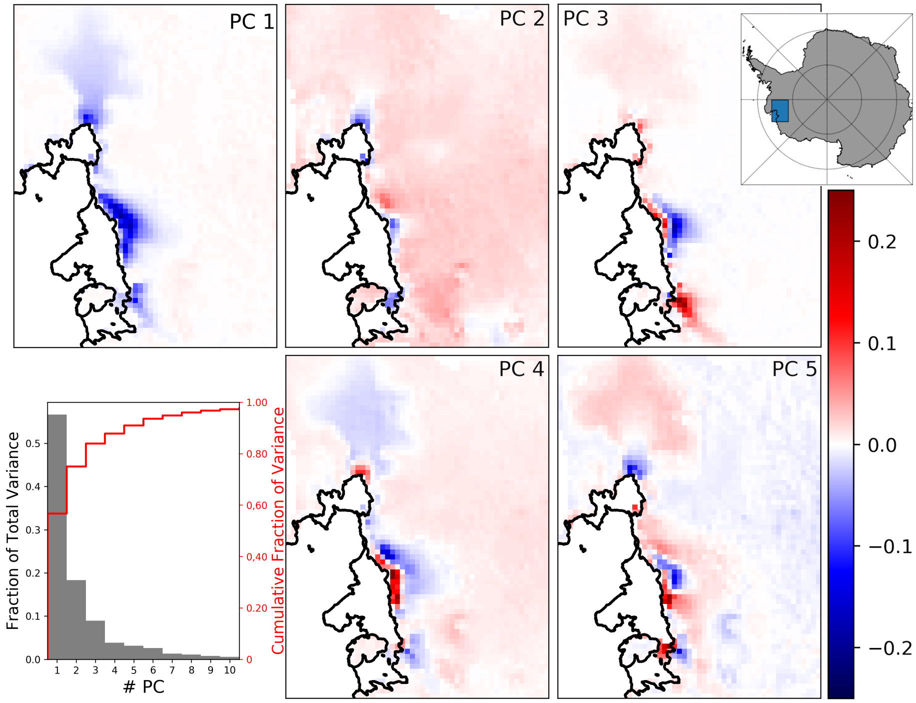


**Figure 2.** The first five normalized PCs of the model ice thickness change fields, building an orthogonal basis. They represent the main modes of variation in the model ensemble and are unitless since normalized. The lower left graph shows the fraction of total variance represented by each PC individually (grey) or sum of the first X PCs (red), based on squared singular values

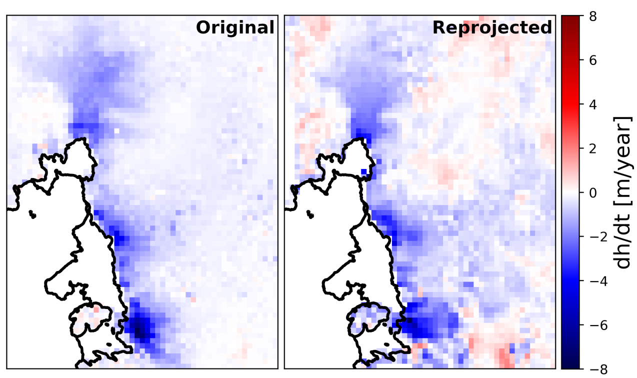


**Figure 3.** Left: Mean observed ice thickness change 2008-2015 based on data from Konrad et al. (2017). Right: ~~Observed ice thickness change~~ as left but projected to first five PCs and ~~reprojected~~ re-projected to spatial field

as repeated observations of the same point in time to specify $\hat{z}$ as the mean and use the variance among the 14 realizations of $\hat{z}$ to define the observational uncertainty in the calibration model (sec 3.4).

Figure 3 shows that large parts of the observations can be represented by the first ~~5~~ five PCs from Fig. 2. ~~It is only this part illustrated on the right of Fig. 3 which is used for calibration. The~~ This is supported by the fact that the spatial variance of the difference between the reprojected and original fields is substantially smaller than from $\boldsymbol{z}_{(\boldsymbol{xy})}$ alone:

$$\frac{VAR(\boldsymbol{z}_{(\boldsymbol{xy})} - \mathbf{B}'((\mathbf{B}'^T\mathbf{B}')^{-1}\mathbf{B}'^T\boldsymbol{z}_{(\boldsymbol{xy})}))}{VAR(\boldsymbol{z}_{(\boldsymbol{xy})})} \approx 0.58$$

It is only the part of the observations which can be represented by five PCs (right of Fig. 3) which will influence the calibration.

### 3.3 Emulation

For ~~probabilistic projections~~ a probabilistic assessment we need to consider the probability density in the full, five-dimensional parameter space. This exploration can require very dense sampling of probabilities in the input space to ensure appropriate representation of all probable parameter combinations. This is especially the case if the calibration is favouring only small subsets of the original input space. In our case more than 90% of the ~~projection~~ calibrated distribution would be based on just five BISICLES ensemble members. For computationally expensive models sufficient sampling can be achieved by statistical emulation, as laid out in the following.

A row of $\mathbf{V}'^T$ can be understood as indices of how much of a particular principal component is present in every ice sheet model simulation. Emulation is done by replacing the discrete number of ice sheet model simulations by continuous functions or statistical models. We use each row of $\mathbf{V}'^T$, combined with $\boldsymbol{\Theta}$, to train an independent statistical model where the mean of the random distribution at $\boldsymbol{\theta}$ is denoted $\omega_i(\boldsymbol{\theta})$. Here the training points are noise free as the emulator is representing a deterministic ice sheet model and therefore $\omega_i(\boldsymbol{\Theta}) = [\mathbf{V}'^T]_i$ for principal components $i = 1,...,k$. Each of those models can be used to interpolate (extrapolation should be avoided) between members of $\boldsymbol{\Theta}$ to predict the ice sheet model behaviour and create surrogate ensemble members.

We use Gaussian Process (GP) models, which are a common choice for their high level of flexibility and inherent emulation uncertainty representation (Kennedy and O'Hagan, 2001; O'Hagan, 2006; Higdon et al., 2008). The random distribution of a Gaussian process model with noise free training data at a new set of input values $\boldsymbol{\theta}_*$ is found by (e.g. Rasmussen and Williams, 2006):

$$\begin{aligned}\Omega_{i*} = N(\,&K(\boldsymbol{\theta}_*,\boldsymbol{\Theta})K(\boldsymbol{\Theta},\boldsymbol{\Theta})^{-1}\omega_i(\boldsymbol{\Theta}),\\ &K(\boldsymbol{\theta}_*,\boldsymbol{\theta}_*) - K(\boldsymbol{\theta}_*,\boldsymbol{\Theta})K(\boldsymbol{\Theta},\boldsymbol{\Theta})^{-1}K(\boldsymbol{\Theta},\boldsymbol{\theta}_*))\end{aligned} \tag{4}$$

where $N(\cdot,\cdot)$ represents a multivariate normal distribution and the values of $K(\boldsymbol{\Theta},\boldsymbol{\Theta})_{ij} = c(\boldsymbol{\theta_i},\boldsymbol{\theta_j})$ are derived from evaluations of the GP covariance function $c(\cdot,\cdot)$. Equivalent definitions are used for $K(\boldsymbol{\theta}_*,\boldsymbol{\Theta})$, $K(\boldsymbol{\Theta},\boldsymbol{\theta}_*)$ and $K(\boldsymbol{\theta}_*,\boldsymbol{\theta}_*)$, note that $K(\boldsymbol{\theta}_*,\boldsymbol{\theta}_*)$ is a $1\times1$ matrix if we emulate one new input set at a time. We use a Matern ($\frac{5}{2}$) type function for $c(\cdot,\cdot)$ which describes the covariance based on the distance between input parameters. Coefficients for $c(\cdot,\cdot)$ (also called hyper-parameters), including the correlation length scale, are optimized on the marginal likelihood of $\omega(\boldsymbol{\Theta})$ given the GP. We refer to Rasmussen and Williams (2006) for an in-depth discussion and tutorial of Gaussian Process Emulators.

Due to the statistical independence of the principal components we can combine the $k$ GPs to:

$$\mathbf{\Omega} = N(\boldsymbol{\omega}(\boldsymbol{\theta}), \quad \mathbf{\Sigma}_\omega(\boldsymbol{\theta})) \tag{5}$$

The combined $\mathbf{\Omega}$ is in the following called emulator and $\boldsymbol{\omega}(\boldsymbol{\theta})$ as well as the entries of the diagonal matrix $\mathbf{\Sigma}_\omega(\boldsymbol{\theta})$ follow from ~~equation 4~~Eq. (4). We use the python module GPy for training (GPRegression()) and marginal likelihood optimization (optimize_restarts()). In total we generate more than 119 000 emulated ensemble members. Emulator estimates of ice sheet model values in a leave-one-out cross-validation scheme are very precise with squared correlation coefficients for both emulators of $R^2 > 0.988$ (~~See~~ see supplement for more information).

### 3.4 Calibration Model

Given the emulator in basis representation, a calibration can be performed either after re-projecting the emulator output back to the original spatial field (e.g. Chang et al., 2016a; Salter et al., 2018) or in the basis representation itself (e.g. Higdon et al., 2008; Chang et al., 2014). Here we will focus on the PC basis representation.

We assume the existence of a parameter configuration $\boldsymbol{\theta}^*$ within the bounds of $\mathbf{\Theta}$ (the investigated input space) which leads to an optimal model representation of the real world. To infer the probability of any $\boldsymbol{\theta}$ to be $\boldsymbol{\theta}^*$ we rely on the existence of observables, i.e. model quantities $\boldsymbol{z}$ for which corresponding measurements $\hat{\boldsymbol{z}}$ are available. We follow Bayes' theorem to update prior (uninformed) expectations about the optimal parameter configuration with the observations to find posterior (updated) estimates. The posterior probability of $\boldsymbol{\theta}$ being the optimal $\boldsymbol{\theta}^*$ given the observations is:

$$\pi(\boldsymbol{\theta}|\boldsymbol{z} = \hat{\boldsymbol{z}}) \propto L(\boldsymbol{z} = \hat{\boldsymbol{z}}|\boldsymbol{\theta}) \times \pi(\boldsymbol{\theta}) \tag{6}$$

where $L(\boldsymbol{z} = \hat{\boldsymbol{z}}|\boldsymbol{\theta})$ is the likelihood of the observables to be as they have been observed under the condition that $\boldsymbol{\theta}$ is $\boldsymbol{\theta}^*$, and $\pi(\boldsymbol{\theta})$ is the prior (uninformed) probability that $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Following Nias et al. (2016) we choose uniform prior distributions in the scaled parameter range [0,1] (see also section 2 and Eq. 11 in Nias et al. (2016)). The emulator output is related to the real state of the ice sheets in basis representation, $\boldsymbol{\gamma}$, by the model discrepancy $\boldsymbol{\varepsilon}$:

$$\boldsymbol{\gamma} = \boldsymbol{\omega}(\boldsymbol{\theta}^*) + \boldsymbol{\varepsilon} \tag{7}$$

We assume the model discrepancy to be multivariate Gaussian distributed with zero mean; $\boldsymbol{\varepsilon} = N(0, \quad \mathbf{\Sigma}_\varepsilon)$. The observables are in turn related to $\boldsymbol{\gamma}$ by:

$$\boldsymbol{z} = \boldsymbol{\gamma} + (\mathbf{B}'^T\mathbf{B}')^{-\mathbf{1}}\mathbf{B}'^T\boldsymbol{e} \tag{8}$$

where $\boldsymbol{e}$ is the spatial observational error and the transformation $(\mathbf{B}'^T\mathbf{B}')^{-\mathbf{1}}\mathbf{B}'^T$ follows from Eq. 3.

We simplify the probabilistic inference by assuming the model error/discrepancy $\boldsymbol{\varepsilon}$, the model parameter values $\mathbf{\Theta}$ and observational error $\boldsymbol{e}$ to be mutually statistically independent and $\boldsymbol{e}$ to be spatially identically distributed with variance $\sigma_e^2$, so that

$$(\mathbf{B}'^T\mathbf{B}')^{-\mathbf{1}}\mathbf{B}'^T\boldsymbol{e} = N(0, \quad \sigma_e^2(\mathbf{B}'^T\mathbf{B}')^{-\mathbf{1}}) \tag{9}$$

The $k \times k$ matrix $(\mathbf{B'}^T\mathbf{B'})^{-1}$ is diagonal with the element-wise inverse of $diag(\mathbf{S'})_i^2$ as diagonal values. We estimate $\sigma_e^2$ from the variance among the 14 observational periods for the first principal component constituting $\hat{z_1}$, i.e.

$$\sigma_e^2 = VAR(\hat{z_1}) \cdot diag(\mathbf{S'})_1^2 \tag{10}$$

Note that the existence of $\boldsymbol{\gamma}$ is an abstract concept, implying that it is only because of an error $\varepsilon$ that we cannot create a numerical model which is equivalent to reality. However abstract, it is a useful, hence common statistical concept allowing us to structure expectations of model and observational limitations (Kennedy and O'Hagan, 2001). Neglecting model discrepancy, whether explicitly by setting $\boldsymbol{\varepsilon} = \mathbf{0}$, or implicitly, would imply that an ice sheet model can make exact predictions of the future once the right parameter values are found. This expectation is hard to justify considering the assumptions which are made for the development of ice sheet models, including sub-resolution processes. Neglecting model discrepancy typically results in overconfidence and potentially biased results.

~~At the same time can the~~ The inclusion of model discrepancy can at the same time lead to identifiability issues where the model signal cannot be distinguished from imposed systematic model error. ~~To overcome such issues, constraints on the , e.g. spatial shape ,~~ Constraints on the spatial shape of the discrepancy ~~are used~~ have been used to overcome such issues (Kennedy and O'Hagan, 2001; Higdon et al., 2008). An inherent problem with representing discrepancy is that its amplitude and spatial shape are in general unknown. If the discrepancy were well understood the model itself or its output could be easily corrected. Even if experts can specify regions or patterns which are likely to show inconsistent behaviour, it cannot be assumed that these regions or patterns are the only possible forms of discrepancy. If its representation is too flexible it can however become numerically impossible in the calibration step to differentiate between discrepancy and model behaviour.

For these reasons we choose a rather heuristic method which considers the impact of discrepancy on the calibration directly and independently for each PC. Therefore $\boldsymbol{\Sigma}_\varepsilon$ is diagonal with $diag(\boldsymbol{\Sigma}_\varepsilon) = (\sigma_{\varepsilon 1}^2, ..., \sigma_{\varepsilon k}^2)^T$. The 'three sigma rule' states that at least 95% of continuous unimodal density functions with finite variance lie within three standard deviations from the mean (Pukelsheim, 1994). For the $i$.th PC we therefore find $\sigma_{i95}^2$ so that 95% of the observational distribution $N(\hat{z}_i, \quad \sigma_{ei}^2)$ lies within $3\sigma_{i95}$ from the mean of $\boldsymbol{\omega}(\boldsymbol{\Theta})_i$, i.e. across the $n$ ensemble members. We further note that we do not know the optimal model setup better than we know the real state of the ice sheet and set the minimum discrepancy to the observational uncertainty. Hence $\sigma_{\varepsilon i}^2 = MAX(\sigma_{i95}^2, \sigma_{ei}^2)$.

We thereby force the observations to fulfill the 'three-sigma rule' by considering them as part of the model distribution $\boldsymbol{\omega}(\boldsymbol{\Theta})_i$ while avoiding over confidence in cases where observations and model runs coincide.

### 3.4.1 History matching

Probabilistic calibrations search for the best input parameters, but stand-alone probabilistic calibrations cannot guarantee that those are also 'good' input parameters in an absolute sense. While 'good' is subjective, it is possible to define and rule out implausible input parameters. The Implausibility parameter is commonly defined as (e.g. Salter et al., 2018):

$$\mathcal{I}(\boldsymbol{\theta}) = (\boldsymbol{\omega}(\boldsymbol{\theta}) - \hat{\boldsymbol{z}})^T \boldsymbol{\Sigma}_T^{-1} (\boldsymbol{\omega}(\boldsymbol{\theta}) - \hat{\boldsymbol{z}}) \tag{11}$$

with $\mathbf{\Sigma}_T = \sigma_e^2(\mathbf{B'}^T\mathbf{B'})^{-\mathbf{1}} + \mathbf{\Sigma}_\varepsilon + \mathbf{\Sigma}_\omega$. A threshold on $\mathcal{I}(\boldsymbol{\theta})$ can be found using the 95% interval of a chi-squared distribution with $k = 5$ degrees of freedom. Therefore we rule out all $\boldsymbol{\theta}$ with $\mathcal{I}(\boldsymbol{\theta}) > 11$. By adding this test, called history matching, we ensure that only those input parameters are used for a probabilistic calibration which are reasonably close to the observations. In the worst case the whole input space could be ruled out, forcing the practitioner to reconsider the calibration approach and uncertainty estimates. Here about 1.4% of the parameter space cannot be ruled out.

### 3.4.2 Probabilistic Calibration

For all $\boldsymbol{\theta}$ which have not been ruled out, the likelihood $L(\boldsymbol{z} = \hat{\boldsymbol{z}}|\boldsymbol{\theta})$ follows from ~~equations 5, 8, 7 and 9~~Eq. (5), Eq. (8), Eq. (7) and Eq. (9):

$$L(\boldsymbol{z} = \hat{\boldsymbol{z}}|\boldsymbol{\theta}) \propto exp\left[-\frac{1}{2}(\boldsymbol{\omega}(\boldsymbol{\theta}) - \hat{\boldsymbol{z}})^T \mathbf{\Sigma}_T^{-1}(\boldsymbol{\omega}(\boldsymbol{\theta}) - \hat{\boldsymbol{z}})\right] \tag{12}$$

The calibration distribution in ~~Equation 6~~Eq. (6) can be evaluated using Eq. 12 with a trained emulator (Eq. 4), observational (Eq. 10) and model discrepancy (above) and the prior parameter distributions $\pi(\theta)$ set by expert judgment.

## 3.5 Calibration model test

In this section we test our calibration approach on synthetic observations to see whether our method is capable of finding known-correct parameter values. We select one member of the BISICLES model ensemble at a time and add 14 different realizations of noise to it. The noise is added to see how the calibration performs if the observations cannot be fully represented by the ice sheet model.

We use spatially independent, zero-mean, normally distributed, random noise with variance equal to the local variance from the 14 periods of satellite observations. This way the variance incorporates dynamic changes (acceleration/deceleration of the ice thickness change) and technical errors (e.g. measurement and sampling errors). For each selected model run we generate 14 noise fields and add them to the single model ice thickness change field. These 14 realizations ~~are used in exactly the same way as described before for~~ replace the 14 periods of satellite observations for the synthetic model tests.

For ~~Figure~~ Fig. 4 the model run with central parameter values ($= 0.5$) for basal traction, viscosity and ocean melt scaling factors, nonlinear ~~sliding~~ friction and modified bedrock has been selected, as indicated by black circles. This parameter set has been selected as it highlights the limitations of the calibration, ~~but~~ the results of ~~many~~ eleven other synthetic model tests are shown in the supplement.

Figure 4 illustrates which parts of the model input space are most successful in reproducing the synthetic observations of ~~surface elevation~~ ice thickness changes during the calibration period. For visualisation we collapse the five dimensional space onto each combination of two parameters and show how they interact. For a likely (yellow) area in Fig. 4 it is not possible to see directly what values the other three parameters have, but very unlikely (black) areas indicate that no combination of the remaining parameter values results in ~~a good model configuration~~model configurations consistent with observations.

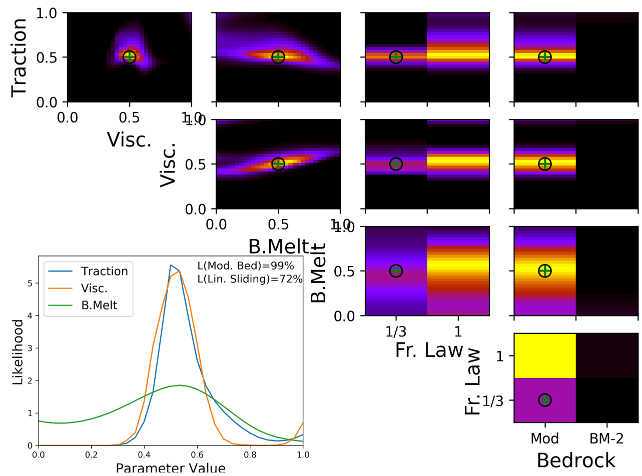


**Figure 4.** Likelihood of parameter combinations of synthetic test case (evaluations of ~~Equation~~ Eq. (12)). Upper right panels show likelihood values marginalized to pairs of parameters, normalized to the respective maximum for clarity. Lower left panel shows likelihood values marginalized to individual parameters for the three scalar parameters (line plots), and ~~sliding~~ friction law and bedrock topography map (text and quotation within), normalized to an integral of one, consistent with Probability Density Functions. The central values for traction, viscosity and ocean melt as well as nonlinear ~~sliding~~ friction and modified bedrock are used. The parameter values are also shown by the black circles, while the values of the set of parameters with highest likelihood are shown by green crosses.

As can be seen from ~~Figure~~ Fig. 4, marginal likelihoods of our calibration approach can favour linear ~~sliding~~ friction even if the synthetic observations use nonlinear ~~sliding~~friction. In addition, the ocean melt parameter is often weakly constrained or, as in this case, biased towards small melt factors. In contrast, the basal traction coefficient and viscosity scaling factors have a strong mode at, or close to, the correct value of 0.5 and the correct bedrock map can always be identified (~~Figure~~ Fig. 4 and supplement). Different values of basal traction and viscosity have been tested in combination with both bedrock maps and show similar performance (see supplement). The fact that the parameter setup used for the test is attributed the maximal likelihood (green cross on top of black circle) supports our confidence in the implementation as the real parameter set is identified correctly as best fit. Relative ambiguity with respect to ~~sliding~~ friction law and ocean melt overrules the weak constraints on these parameters in the marginalized likelihoods. The higher total likelihood of linear ~~sliding~~ friction can be traced back to a higher density of central ensemble members for linear ~~sliding. Nonlinear sliding~~ friction. Nonlinear friction produces more extreme ice sheet simulations as ~~fast simulations~~ simulations with high velocities will have reduced (compared to linear ~~sliding~~friction) basal drag and ~~become even faster~~ speed up even more (and vice versa for ~~slow simulations~~ simulations with slow ice flows). The frequency distribution of total sea level contribution and basis representation are therefore wider for nonlinear ~~sliding (~~friction (see supplement). The relative density of ensemble members around the mode of the frequency distribution can, as for this test case, cause a smaller marginal likelihood for nonlinear ~~sliding~~ friction compared to linear ~~sliding~~ friction (28% to 72%). This can be considered a caveat of the model ensemble which might very well be present in other ensembles which perturb the friction law in combination with other parameters. If the friction law cannot be adequately constrained, as is the case for all calibration approaches tested here, the prior believe in the optimal friction law must be set very carefully.

~~But why is the signal of sliding~~

The signal of friction law and ocean melt is not strong enough to adequately constrain the calibration, even though both parameters are known ~~(Arthern and Williams, 2017; Joughin et al., 2019)~~ to have a strong impact on ~~model simulations?~~ the ice sheet (Pritchard et al., 2012; Arthern and Williams, 2017; Jenkins et al., 2018; Joughin et al., 2019; Brondex et al., 2019). This is likely related to the ~~delayed~~ slower impact of those parameters compared to the others. A change in bedrock, basal traction or viscosity has a much more immediate effect on the ice dynamics. For example, if the basal traction field is halved, the basal drag will be reduced by the same amount leading to a speed up of the ice at the next time step (via the solution of the stress balance). The perturbation of ocean melt from the start of the model period has to significantly change the ice shelf thickness before the ice dynamics upstream are affected. The ~~fields of basal traction coefficient are adjusted to the sliding law by the inversion of surface ice velocities so~~ initialization of the ensemble has been performed for each friction law individually which means that the initial ~~basal drag $\tau_b$ is approximately the same for both sliding laws with:~~

$$\tau_b = C_m(x,y) \cdot |v(x,y,t)|^{m-1} \cdot v(x,y,t)$$

~~where $C_m(x,y)$ is the spatial basal traction coefficient for sliding law exponent m ($m = 1$ for linear, $m = 1/3$ for nonlinear sliding) and $v(x,y,t)$ being the basal ice velocity. As $C_m(x,y)$ compensates for $|v(x,y,t)|^{m-1}$ at the beginning of the model period, it~~ speed of the ice is by design equivalent. It is only after the ice velocities change that the ~~sliding~~ different degrees of linearity in the friction law has any impact on the simulations. ~~A change in bedrock, basal traction or viscosity have , however, a~~ This does not mean that the simulations are insensitive to the ocean melt forcing and friction law, in fact Fig. 4 shows that both parameters have some impact on the simulation in the calibration period. It just means that the much more immediate ~~effect on the ice dynamics and are therefore expected~~ effects of basal traction and viscosity are likely to dominate the calibration on short time scales.

From this test we conclude that basal ~~sliding~~ friction law and ocean melt scaling cannot be inferred ~~from~~ with this calibration approach and calibration period. We will therefore only calibrate the bedrock as well as basal traction and viscosity scaling factors. Several studies used the observed dynamical changes of parts of the ASE to test different ~~sliding~~ friction laws. Gillet-Chaulet et al. (2016) find a better fit to evolving changes of Pine Island Glacier surface velocities for smaller m, reaching a minimum of the cost function from around m=1/5 and smaller. This is supported by Joughin et al. (2019) who find m=1/8 to capture the PIG speed up from 2002 to 2017 very well, matched only by a regularized Coulomb (Schoof-) ~~sliding~~ friction law. It further is understood, that parts of the ASE bed consist of sediment-free, bare rocks for which a linear Weertman ~~sliding~~ friction law is not appropriate (Joughin et al., 2009). We therefore select nonlinear ~~sliding~~ friction by expert judgment and use a uniform prior for the ocean melt scaling.

### 3.6 Comparison with other calibration approaches

To put the likelihood distribution from ~~Figure~~ Fig. 4 into context, we try two other methodical choices. First we calibrate in the spatial domain after re-projecting from the emulator results.

$$\boldsymbol{y'}(\boldsymbol{\theta}) = \mathbf{B}'\boldsymbol{\omega}(\boldsymbol{\theta}) \tag{13}$$

where $\boldsymbol{y'}(\boldsymbol{\theta_i})$ are the re-projected ice sheet model results after truncation for parameter setup $\boldsymbol{\theta}$. We set the model discrepancy to twice the observational uncertainty $\sigma_e^2$ so that the re-projected likelihood $L_{(xy)}$ simplifies to:

$$L_{(xy)}(\boldsymbol{z_{(xy)}}|\boldsymbol{\theta}) \propto \prod_{i=1}^{m} exp\left[-\frac{1}{2}\frac{(y'(\boldsymbol{\theta})_i - z_{(xy)i})^2}{3\sigma_e^2}\right] \tag{14}$$

Another approach is to use the net yearly sea level contribution from the observations $SLC(\boldsymbol{z_{(xy)}})$ and model $SLC(\boldsymbol{y'}(\boldsymbol{\theta_i}))$ for calibration, as done in e.g. Ritz et al. (2015).

$$L_{SLC}(\boldsymbol{z_{(xy)}}|\boldsymbol{\theta}) \propto exp\left[-\frac{1}{2}\frac{(SLC(\boldsymbol{y'}(\boldsymbol{\theta})) - SLC(\boldsymbol{z_{(xy)}}))^2}{3\sigma_{SLC}^2}\right] \tag{15}$$

Again, we set the model discrepancy to twice the observational uncertainty which we find from the variance of the yearly sea level contributions for the 14 bi-yearly satellite intervals. $\sigma_{SLC}^2 = VAR(SLC(\boldsymbol{z_{(xy)}})) = 0.035^2[\frac{mmSLE^2}{year^2}]$.

## 4 Results

~~The~~ Results for the synthetic model test for the calibration in (x,y) representation (~~Figure~~ Fig. 5a) ~~behaves similarly to the~~ show similar behavior as for basis representation (~~Figure~~ Fig. 4) in that ~~sliding~~ friction law exponent and, to a lesser degree, basal melt are weakly constrained while the confidence in the correctly identified traction and viscosity values is even higher. Using only the net sea level rise contribution constrains the parameters weakly; it shares the limitations of not constraining the ocean melt and favouring linear ~~sliding~~ friction but in addition, a wide range of traction-viscosity combinations perform equally well and there is no constraint on bedrock (~~Figure~~ Fig. 5b). Furthermore, the model run used as synthetic observations is not identified as the most likely setup in ~~Figure~~ Fig. 5b. This demonstrates the value of the extra information - and stronger parameter constraints - provided by the use of two-dimensional observations.

## 5 ~~Results~~

~~Following the synthetic model test, we now calibrate traction, viscosity and bedrock with the satellite data.~~

~~The calibration~~ Moving on to using satellite data, the basis-calibration finds that the modified bedrock from Nias et al. (2016) produces much more realistic ~~surface elevation~~ ice thickness changes than the original Bedmap2 topography (Fig. 6a). The weighted average of basal traction and ~~velocity~~ viscosity parameters are 0.47 and 0.45, respectively, which is slightly smaller than the default values (0.5). This amounts to a 3.5% and 7.2% reduction in amplitude compared to the optimized fields

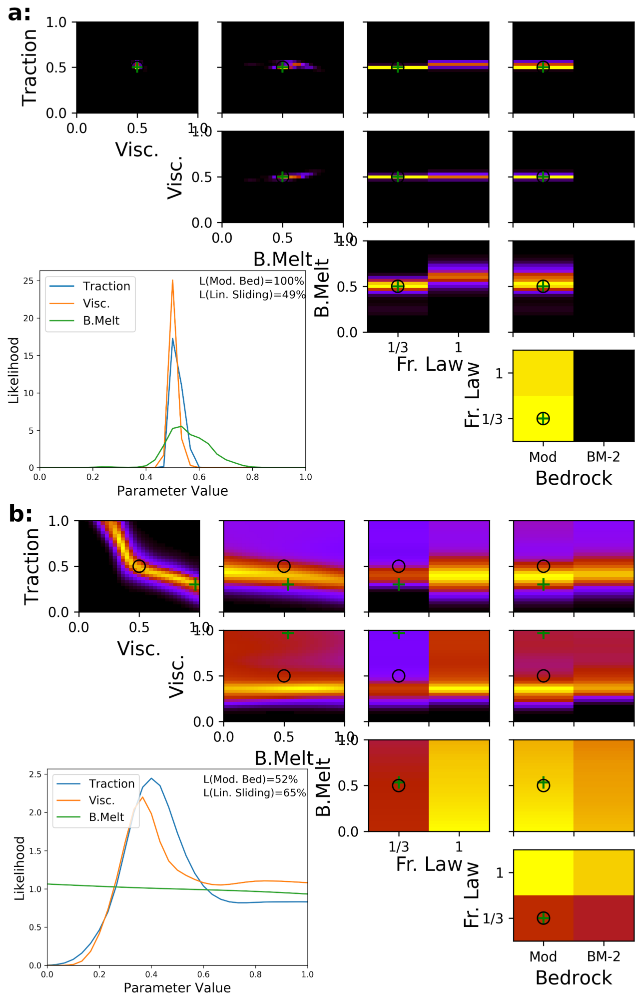


**Figure 5.** Likelihood of parameter combinations of synthetic test case for reprojected emulator estimates (top, a; Equation (14)) and sea level rise contribution calibration (bottom, b; Equation (15)). Upper right panels show likelihood values marginalized to pairs of parameters, normalized to the respective maximum for clarity. Lower left panel shows likelihood values marginalized to individual parameters for the three scalar parameters (line plots), and ~~sliding~~ friction law and bedrock topography map (text and quotation within), normalized to an integral of one, consistent with Probability Density Functions. The central values for traction, viscosity and ocean melt as well as nonlinear ~~sliding~~ friction and modified bedrock are used. The parameter values are also shown by the black circles, while the values of the set of parameters with highest likelihood are shown by green crosses.

~~from~~ by (Nias et al., 2016). While this reduction is relatively small and the central run cannot be ruled out as optimal setup (its likelihood to be optimal is notably larger than zero), this does indicate a possible underestimation of sea level contribution by the default run. With modified bedrock, non-linear friction law and default traction and viscosity values, the SLCs at the end

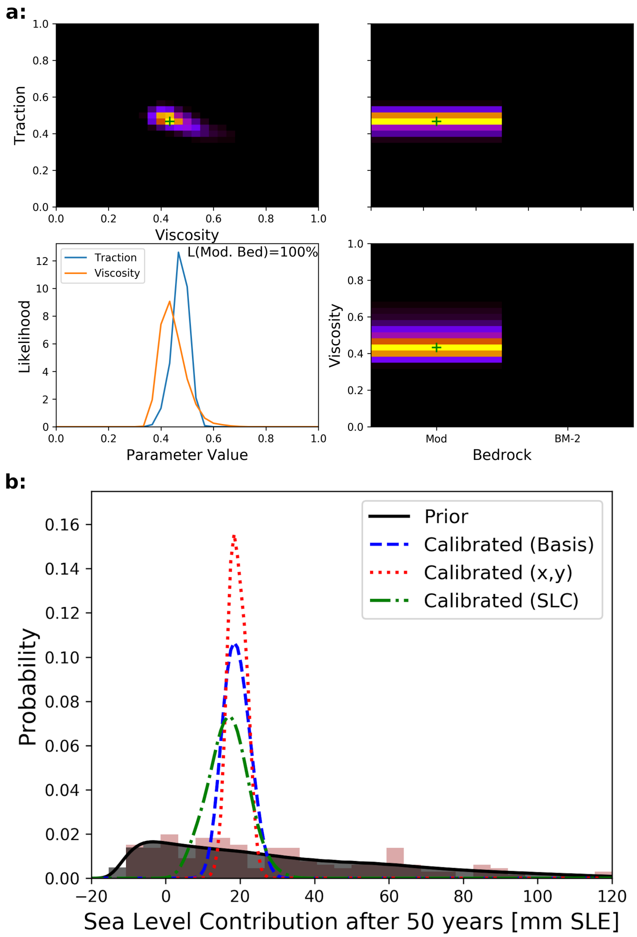


**Figure 6.** a: Likelihood of parameter combinations in basis representation from satellite observations (evaluations of ~~Equation~~ Eq. (12)). Upper right panels show likelihood values marginalized to pairs of parameters, normalized to the respective maximum for clarity. Lower left panel shows likelihood values marginalized to individual parameters for the two scalar parameters (line plots) and bedrock topography map (text and quotation within), normalized to an integral of one in the style of Probability Density Functions. Values of the set of parameters with highest likelihood are shown by green crosses. b: Projected sea level rise contributions at the end of model period for uncalibrated BISICLES runs (brown shades), uncalibrated emulator calls (Grey shade) and different calibration approaches (colored lines).

of the simulation period range from 11 to 19.5 mm SLE depending on the ocean melt scaling, while the basis-calibration mean SLC is 19.1 mm SLE (Table 1).

~~We~~ For updated probability distributions of sea level contribution after 50 years in Fig. 6b we use the calibration in basis representation (likelihood shown in Fig. 6a) as well as the reprojected (x,y) and SLC based calibrations~~to update the projections of~~ . The three calibration approaches are consistent (large overlap) while using the re-projection approach leads to the most narrow SLC distribution (Fig. 6b), as was indicated by the findings of Section 3.6. Calibration on the total sea level contribution leads to a wider distribution with the lower bound (5 %-ile) being more than 6 mm SLE smaller than for the two other approaches. All

**Table 1.** Total sea level contribution after 50 years in mm SLE: (weighted) mean, most likely contribution and percentiles; with and without calibrations.

| | Mean | Mode | 5% | 25% | 50% | 75% | 95% |
|---|---|---|---|---|---|---|---|
| Prior | 30.6 | -3.3 | -8.4 | 4.2 | 23.1 | 51.3 | 94.5 |
| Posterior basis | 19.1 | 18.4 | 13.9 | 16.7 | 18.9 | 21.4 | 24.8 |
| Posterior (x,y) | 19.2 | 18.4 | 16.7 | 17.7 | 18.6 | 21.1 | 22.2 |
| Posterior SLC | 16.8 | 17.5 | 7.7 | 13.2 | 16.8 | 20.3 | 25.6 |

of them strongly reduce uncertainties compared to the uncalibrated prior distribution with the 90% confidence interval width reducing to 10.9 mm SLE (basis-calibration), 5.5 mm SLE (reprojected-calibration) and 17.9 mm SLE (SLC-calibration) from 102.9 mm SLE (uncalibrated) (Fig. 6b and ~~grounding line retreat after 50 years in Figure 6b . As can be seen from the Grey and Brown shaded histograms in~~ Table 1). Figure 6b ~~(emulated and original BISICLES ensemble)~~ also shows histograms of the emulated and the original BISICLES ensembles (grey and brown shades) and illustrates how the emulation helps to overcome challenges of limited sample size.

~~The three calibration approaches are consistent (large overlap) wile using the reprojection approach leads to the most narrow SLC distribution (Figure 6b), as was indicated by the findings of Section 3.6. Calibration on the total sea level contribution leads to a wider distribution with the lower bound of projections (5 %-ile) being more than 6 mmSLE smaller than for the two other approaches. All of them strongly reduce projection uncertainties compared to the uncalibrated prior distribution (Figure 6b and Table 1)~~

## 5 Discussion

In general, previous Antarctic ice sheet model uncertainty studies have either focused on parameter inference (Chang et al., 2016a, b; Pollard et al., 2016), or made projections that are not calibrated with observations ~~(Nias et al., 2016; Schlegel et al., 2018; Bulthuis~~ with the remaining probabilistic calibrated projections being based on simple (fast) models using highly aggregated observations and some relying heavily on expert judgment (Ruckert et al., 2017; Ritz et al., 2015; Little et al., 2013; Levermann et al., 2014; DeConto and Pollard, 2016; Edwards et al., 2019). Here we perform statistically-founded parameter inference using spatial observations to calibrate high resolution, grounding line resolving ice sheet model ~~projections~~simulations.

The theoretical basis for most of the methodology used here has been laid out in Higdon et al. (2008), including the Principal Component (PC) decomposition, emulation and model calibration in the PC space. This calibration in basis representation has been adapted and tested for general circulation (climate) and ocean models (Sexton et al., 2012; Chang et al., 2014; Salter et al., 2018; Salte By combining this approach with a simple but robust discrepancy representation, we attempt to bridge the gap between the demanding mathematical basis and practical applications in geoscience. We compare a novel calibration of a grounding line resolving ice sheet model in the PC space with a reprojected calibration which assumes that the difference between observations and calibration model are spatially uncorrelated (like e.g. Chang et al., 2016b). In comparison with studies that calibrate the total sea level contribution (like e.g. Ritz et al., 2015), we are able to exploit more of the available observational information to add further constraints to the input parameters and sharpen the posterior distribution (Fig. 5 and 6b). Similar improvements should be achievable for ice sheet simulations forced by global climate model projections.

The modified bedrock removes a topographic rise near the initial grounding line of Pine Island Glacier which could be caused by erroneous observations (Rignot et al., 2014). This rise, if present, would have a stabilizing effect on the grounding line and simulations without it can result in more than twice the sea level contribution from Pine Island Glacier for some ~~sliding~~ friction laws (Nias et al., 2018). Here we find the modified bedrock topography to produce a spatial response far more consistent with observed ~~surface elevation~~ ice thickness changes than for the original Bedmap2 bedrock (Fig. 6a). The modified bedrock has been derived by reducing clearly unrealistic behaviour of the same ice sheet model, a better calibration performance was therefore to be expected. However, no satellite observations have been used for the bedrock modification in Nias et al. (2016), nor has there been a quantitative probabilistic assessment.

The non-spatial calibration on total sea level contribution alone cannot distinguish between the two bedrocks (~~Figure~~ Fig. 5b). ~~Projections~~ Simulations for this region based on Bedmap2, calibrated on the SLC are likely to either be compensating the overly-stabilising bedrock with underestimated viscosity and/or traction coefficients, or underestimating the sea level contribution altogether. In addition to the unconstrained bedrock, the SLC calibration permits a wide range of traction and viscosity coefficients, including values far from the correct test values (~~Figure~~ Fig. 5b). This shows that the SLC calibration permits more model runs which are right for the wrong reasons; they have approximately the right sea level rise contribution in the calibration period but can still be poor representations of the current state of the ice sheet.

The extremely small area of likely input parameters for the reprojected (x,y) calibration (~~Figure~~ Fig. 5a and Supplement) could indicate overconfidence in the retrieved parameter values, but could also mean that the available information is exploited more efficiently. Using subsections of the calibration period has a small impact on basis and SLC ~~calibration~~calibrations. However, for one of the sub-periods with ~~reprojected~~ re-projected calibration the probability interval does not overlap with the results of the whole ~~7~~ seven year calibration period (Table 1 in the Supplement). Since the sub-period is part of the ~~7~~ seven year period we would expect the results to be non-contradictory, indicating that the probability intervals are too narrow and hence the approach, as implemented here, being overconfident. The different ways of handling model discrepancy influence the width of the probability intervals.

The average sea level contribution from the observations used here is 0.36 mm SLE per year, consistent with estimates form McMillan et al. (2014) of ~~$0.33 \pm 0.05$~~ $0.33 \pm 0.05$ mm SLE per year for the Amundsen Sea Embayment from 2010-2013. Calibrated rates in the beginning of the model period are very similar (0.335, 0.327 and 0.363 mm SLE per year for basis, (x,y) and SLC calibration, respectively). For (x,y) and basis calibration the rates increase over the 50 year period while the rate of mass loss reduces for the SLC calibration (50 year average SLC rates: 0.382, 0.384 and 0.336 mm SLE per year for basis, (x,y) and SLC ~~,~~calibration, respectively). The fact that the SLC calibration starts with the largest rates of sea level contribution but is the only approach seeing a reduction in those rates, in combination with the above mentioned suspicion of it allowing unrealistic setups, raises questions about how reliable calibrations on total sea level contribution alone are.

The ice sheet model data used here is not based on a specific climate scenario but instead projects the state of the ice sheet under current conditions into the future (with imposed perturbations). Holland et al. (2019) suggest a link between anthropogenic greenhouse gas emissions and increased upwelling of warm circumpolar deep water, facilitating melt at the base of Amundsen sea ice shelves. This would imply a positive, climate scenario dependent trend of ocean melt for the model period, superimposed by strong decadal variability ~~(Holland et al., 2019; Jenkins et al., 2016)~~(Holland et al., 2019; Jenkins et al., 2016, 2018). Warmer ocean and air temperatures would enhance melt and accelerate the dynamic response. Neither do the ~~used simulations~~ simulations used here carry the countervailing predicted increase of surface accumulation in a warmer climate (Lenaerts et al., 2016). Edwards et al. (2019) and Golledge et al. (2019) find that the Antarctic ice sheet response to very different greenhouse gas emissions scenarios starts to diverge from around 2060-2070, while Yu et al. (2018) find ocean melt to have a negligible impact for the first 30 years for their simulations of Thwaites glacier. Combined, this is indicating that climate scenarios would have a small net impact on ~~our~~ 50-year ~~projections~~simulations.

Relating climate scenarios to local ice shelf melt rates is associated with deep uncertainties itself. CMIP5 climate models are inconsistent in predicting Antarctic shelf water temperatures so that the model choice can make a substantial (>50%) difference in the increase of ocean melt by 2100 for the ASE (Naughten et al., 2018). 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 (Favier et al., 2019). The location of ocean melt can be as important as the integrated melt of an ice shelf (Goldberg et al., 2019). The treatment of melt on partially floating grid cells further impacts ice sheet models significantly, even for fine spatial resolutions of 300 m (Yu et al., 2018). It is therefore very challenging to make robust climate scenario dependent ice sheet model predictions. Instead we use ~~projections~~ simulations 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. The method presented here can be applied to forced simulations which would benefit from reduced uncertainty intervals to highlight the impact of climate change on ice sheet models.

The truncation of a principal component decomposition can cause or worsen problems related to the observations not being in the analyzed model output space (see difference in ~~Figure~~Fig. 3). This can mean that there is no parameter configuration $\boldsymbol{\theta}$ which is a good representation of the observations. Basis rotations have been proposed to reduce this problem (Salter et al., 2018); however, here we use only the portions of the observations which can be represented in the reduced PC space (~~Figure~~ Fig. 3b) and argue that configurations which are able to reproduce those portions are likely to be better general representations than those configurations which cannot. We further include a discrepancy variance for each PC to account for systematic observation-model differences, including PC truncation effects and perform an initial history matching to ensure the observations are reasonable close to model results.

The model perturbation has been done by amplitude scaling of the optimized input fields alone, other types of variations to the ~~input~~ basal traction coefficient fields could potentially produce model setups with better agreement to the observations (Petra et al., 2014; Isaac et al., 2015). However, computational and methodological challenges make simple scaling approaches more feasible and the use of a published dataset bars us from testing additional types of perturbations. ~~Probabilistic calibrations~~

~~are an assessment of model setups to be the best of all tested cases. It has to be clear that this is, despite emulation, a vast simplification in searching for the best of all possible modelsetups imaginable.~~

~~The theoretical basis for most of the methodology used here has been laid out in Higdon et al. (2008), including the principal component decomposition, emulation and model calibration in the PC space. This calibration in basis representation has been adapted and tested for general circulation (climate)and ocean models (Sexton et al., 2012; Chang et al., 2014; Salter et al., 2018; Salter and By combining this approach with a simple but robust discrepancy representation, we attempt to bridge the gap between the demanding mathematical basis and practical applications in geoscience. We compare a novel calibration of a grounding line resolving ice sheet model in the PC space with a reprojected calibration which assumes that the difference between observations and calibration model are spatially uncorrelated (like e.g. Chang et al., 2016b). In comparison with studies that calibrate the total sea level contribution (like e.g. Ritz et al., 2015), we are able to exploit more of the available observational information to add further constraints to the input parameters and sharpen the posterior distribution (Figure 5 and 6b).~~ Emulation helps to improve the sampling of the scaling parameters but does not change the fact that we cannot asses the quality of types of perturbation which are not covered by the ice sheet model.

It should also be noted that for a given ice geometry the surface speed (used for initialisation) and ice thickness change (used for calibration) are not fully independent (conversation of mass). Finding the unperturbed traction and viscosity fields to show good agreement with ice thickness change observations is not surprising, yet a good test of the initialisation process, initialisation data and the quality of the initial ice geometry. For the same reasons, the optimized fields cannot be considered without uncertainty. This uncertainty can be quantified by Ice thickness change observations, as has been shown here. A combined temporal and spatial calibration could help to use even more of the available information captured by observations in regions like the ASE where dynamic changes in the ice sheet ~~have been observed~~took place within the observation period. The temporal component could in particular help to constrain the basal ~~sliding~~ friction law exponent and ocean melt scaling.

## 6 Conclusions

We present probabilistic estimates of the dynamic contribution to sea level of unforced 50 year simulations of the Amundsen Sea Embayment in West Antarctica ~~over the next 50 years~~ from a grounding line resolving ice sheet model. ~~We performed a~~ The Bayesian calibration of a published ice sheet model ensemble with satellite ~~measurements of surface elevation changes from 1992-2015, using~~ estimates of changes in ice thickness from 2008-2015 involves spatial decomposition to increase the amount of ~~information used~~ available information from the observations and emulation techniques to search the parameter space more thoroughly.

The calibration has been tested on synthetic test cases and can reliably constrain the bedrock, basal traction and ice viscosity amplitudes. Identifying the most successful basal ~~sliding~~ friction law and ocean melt rate is more challenging, ~~probably due to their slow impact on ice sheet behaviour compared to the short calibration timescale~~interference of those parameters could benefit from a temporally resolved calibration approach and a longer calibration period. The use of net sea level contribution alone allows a wide range of parameter setups, which share the initial net mass loss. This ambiguity (weak constraint) also results in ~~relative~~ relatively wide sea level contribution probability distributions. The extra information from the use of two-dimensional calibrations adds stronger parameter ~~constrains~~constraints, showing that this method ~~can be used~~ has the potential to reduce uncertainties in ice sheet model projections. We compare and discuss spatial calibrations in both ~~,~~ basis and reprojected representation.

Using satellite observations we find the modified bedrock topography derived by Nias et al. (2016) to result in a quantitatively far more consistent model representation of the Amundsen Sea Embayment than Bedmap2. ~~Compared to prior estimates, the calibrations lead to a drastic reduction in the projection uncertainty by more than 80%. Within the~~ Imposing no climate forcing, the calibrated 50 year ~~model period the~~ Amundsen Sea Embayment ~~is expected to contribute between~~ simulations contribute 18.4 [13.9~~and~~, 24.8~~mm SLE (~~] mm SLE (most likely value and 90% probability interval) ~~with a most likely global sea level contribution of 18.4 mm SLE~~to global mean sea level. Compared to prior estimates, these calibrated values constitute a drastic reduction in uncertainty by nearly 90%.

*Code availability.* Code can be accessed at https://github.com/Andreas948

*Author contributions.* AW conducted the study with TE, PH and NE giving valuable advice on the study design and IN on the model data processing and interpretation. All authors contributed to the interpretation of the study results. AW prepared the manuscript with contributions from all co-authors.

*Competing interests.* The authors declare that they have no competing interests

*Acknowledgements.* We would like to thank Hannes Konrad for sharing and advising on the satellite observations and Mark Brandon for general advice. We also thank the anonymous reviewers which helped to significantly improve this work.

## References


Arthern, R. J. and Williams, C. R.: The sensitivity of West Antarctica to the submarine melting feedback, Geophysical Research Letters, 44, 2352–2359, https://doi.org/10.1002/2017GL072514, 2017.

Bamber, J. L., Westaway, R. M., Marzeion, B., and Wouters, B.: The land ice contribution to sea level during the satellite era, Environmental Research Letters, 13, 063 008, https://doi.org/10.1088/1748-9326/aac2f0, 2018.

Brondex, J., Gillet-Chaulet, F., and Gagliardini, O.: Sensitivity of centennial mass loss projections of the Amundsen basin to the friction law, The Cryosphere, 13, 177–195, https://doi.org/10.5194/tc-13-177-2019, 2019.

Bulthuis, K., Arnst, M., Sun, S., and Pattyn, F.: Uncertainty quantification of the multi-centennial response of the Antarctic ice sheet to climate change, The Cryosphere, 13, 1349–1380, https://doi.org/10.5194/tc-13-1349-2019, 2019.

Chang, W., Haran, M., Olson, R., Keller, K., et al.: Fast dimension-reduced climate model calibration and the effect of data aggregation, The Annals of Applied Statistics, 8, 649–673, https://doi.org/10.1214/14-AOAS733, 2014.

Chang, W., Haran, M., Applegate, P., and Pollard, D.: Calibrating an ice sheet model using high-dimensional binary spatial data, Journal of the American Statistical Association, 111, 57–72, https://doi.org/10.1080/01621459.2015.1108199, 2016a.

Chang, W., Haran, M., Applegate, P., Pollard, D., et al.: Improving ice sheet model calibration using paleoclimate and modern data, The Annals of Applied Statistics, 10, 2274–2302, https://doi.org/10.1214/16-AOAS979, 2016b.

Church, J. A., Clark, P. U., Cazenave, A., Gregory, J. M., Jevrejeva, S., Levermann, A., Merrifield, M. A., Milne, G. A., Nerem, R. S., Nunn, P. D., et al.: Sea level change, Tech. rep., PM Cambridge University Press, 2013.

Cornford, S. L., Martin, D. F., Graves, D. T., Ranken, D. F., Le Brocq, A. M., Gladstone, R. M., Payne, A. J., Ng, E. G., and Lipscomb, W. H.: Adaptive mesh, finite volume modeling of marine ice sheets, Journal of Computational Physics, 232, 529–549, https://doi.org/10.1016/j.jcp.2012.08.037, 2013.

Cornford, S. L., Martin, D., Payne, A., Ng, E., Le Brocq, A., Gladstone, R., Edwards, T. L., Shannon, S., Agosta, C., Van Den Broeke, M., et al.: Century-scale simulations of the response of the West Antarctic Ice Sheet to a warming climate, https://doi.org/10.5194/tc-9-1579-2015, 2015.

DeConto, R. M. and Pollard, D.: Contribution of Antarctica to past and future sea-level rise, Nature, 531, 591, https://doi.org/10.1038/nature17145, 2016.

Edwards, T. L., Brandon, M. A., Durand, G., Edwards, N. R., Golledge, N. R., Holden, P. B., Nias, I. J., Payne, A. J., Ritz, C., and Wernecke, A.: Revisiting Antarctic ice loss due to marine ice-cliff instability, Nature, 566, 58, https://doi.org/10.1038/s41586-019-0901-4, 2019.

Favier, L., Durand, G., Cornford, S. L., Gudmundsson, G. H., Gagliardini, O., Gillet-Chaulet, F., Zwinger, T., Payne, A., and Le Brocq, A. M.: Retreat of Pine Island Glacier controlled by marine ice-sheet instability, Nature Climate Change, 4, 117, 2014.

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

Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J., Barrand, N., Bell, R., Bianchi, C., Bingham, R., Blankenship, D. D., Casassa, G., et al.: Bedmap2: improved ice bed, surface and thickness datasets for Antarctica, https://doi.org/10.5194/tc-7-375-2013, 2013.

Gillet-Chaulet, F., Durand, G., Gagliardini, O., Mosbeux, C., Mouginot, J., Rémy, F., and Ritz, C.: Assimilation of surface velocities acquired between 1996 and 2010 to constrain the form of the basal friction law under Pine Island Glacier, Geophysical Research Letters, 43, 10–311, https://doi.org/10.1002/2016GL069937, 2016.

Gladstone, R. M., Lee, V., Rougier, J., Payne, A. J., Hellmer, H., Le Brocq, A., Shepherd, A., Edwards, T. L., Gregory, J., and Cornford, S. L.: Calibrated prediction of Pine Island Glacier retreat during the 21st and 22nd centuries with a coupled flowline model, Earth and Planetary Science Letters, 333, 191–199, https://doi.org/10.1016/j.epsl.2012.04.022, 2012.

Goldberg, D., Gourmelen, N., Kimura, S., Millan, R., and Snow, K.: How accurately should we model ice shelf melt rates?, Geophysical Research Letters, 46, 189–199, https://doi.org/10.1029/2018GL080383, 2019.

Golledge, N. R., Keller, E. D., Gomez, N., Naughten, K. A., Bernales, J., Trusel, L. D., and Edwards, T. L.: Global environmental consequences of twenty-first-century ice-sheet melt, Nature, 566, 65, https://doi.org/10.1038/s41586-019-0889-9, 2019.

Habermann, M., Maxwell, D., and Truffer, M.: Reconstruction of basal properties in ice sheets using iterative inverse methods, Journal of Glaciology, 58, 795–808, https://doi.org/10.3189/2012JoG11J168, 2012.

Hein, A. S., Woodward, J., Marrero, S. M., Dunning, S. A., Steig, E. J., Freeman, S. P., Stuart, F. M., Winter, K., Westoby, M. J., and Sugden, D. E.: Evidence for the stability of the West Antarctic Ice Sheet divide for 1.4 million years, Nature communications, 7, 10 325, 2016.

Higdon, D., Gattiker, J., Williams, B., and Rightley, M.: Computer model calibration using high-dimensional output, Journal of the American Statistical Association, 103, 570–583, https://doi.org/10.1198/016214507000000888, 2008.

Holden, P. B., Edwards, N., Oliver, K., Lenton, T., and Wilkinson, R.: A probabilistic calibration of climate sensitivity and terrestrial carbon change in GENIE-1, Climate Dynamics, 35, 785–806, https://doi.org/10.1007/s00382-009-0630-8, 2010.

Holden, P. B., Edwards, N. R., Garthwaite, P. H., and Wilkinson, R. D.: Emulation and interpretation of high-dimensional climate model outputs, Journal of Applied Statistics, 42, 2038–2055, https://doi.org/10.1080/02664763.2015.1016412, 2015.

Holland, P. R., Bracegirdle, T. J., Dutrieux, P., Jenkins, A., and Steig, E. J.: West Antarctic ice loss influenced by internal climate variability and anthropogenic forcing, Nature Geoscience, 12, 718–724, https://doi.org/10.1038/s41561-019-0420-9, 2019.

Isaac, T., Petra, N., Stadler, G., and Ghattas, O.: Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet, Journal of Computational Physics, 296, 348–368, https://doi.org/10.1016/j.jcp.2015.04.047, 2015.

Jenkins, A., Dutrieux, P., Jacobs, S., Steig, E. J., Gudmundsson, G. H., Smith, J., and Heywood, K. J.: Decadal ocean forcing and Antarctic ice sheet response: Lessons from the Amundsen Sea, Oceanography, 29, 106–117, 2016.

Jenkins, A., Shoosmith, D., Dutrieux, P., Jacobs, S., Kim, T. W., Lee, S. H., Ha, H. K., and Stammerjohn, S.: West Antarctic Ice Sheet retreat in the Amundsen Sea driven by decadal oceanic variability, Nature Geoscience, 11, 733–738, 2018.

Joughin, I., Tulaczyk, S., Bamber, J. L., Blankenship, D., Holt, J. W., Scambos, T., and Vaughan, D. G.: Basal conditions for Pine Island and Thwaites Glaciers, West Antarctica, determined using satellite and airborne data, Journal of Glaciology, 55, 245–257, https://doi.org/10.3189/002214309788608705, 2009.

Joughin, I., Smith, B. E., and Schoof, C. G.: Regularized Coulomb friction laws for ice sheet sliding: Application to Pine Island Glacier, Antarctica, Geophysical Research Letters, 46, 4764–4771, https://doi.org/https://doi.org/10.1029/2019GL082526, 2019.

Kennedy, M. C. and O'Hagan, A.: Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63, 425–464, https://doi.org/10.1111/1467-9868.00294, 2001.

Khazendar, A., Rignot, E., Schroeder, D. M., Seroussi, H., Schodlok, M. P., Scheuchl, B., Mouginot, J., Sutterley, T. C., and Velicogna, I.: Rapid submarine ice melting in the grounding zones of ice shelves in West Antarctica, Nature communications, 7, 13 243, https://doi.org/10.1038/ncomms13243, 2016.

Konrad, H., Gilbert, L., Cornford, S. L., Payne, A., Hogg, A., Muir, A., and Shepherd, A.: Uneven onset and pace of ice-dynamical imbalance in the Amundsen Sea Embayment, West Antarctica, Geophysical Research Letters, 44, 910–918, https://doi.org/10.1002/2016GL070733, 2017.

Lenaerts, J. T., Vizcaino, M., Fyke, J., Van Kampenhout, L., and van den Broeke, M. R.: Present-day and future Antarctic ice sheet climate and surface mass balance in the Community Earth System Model, Climate Dynamics, 47, 1367–1381, https://doi.org/10.1007/s00382-015-2907-4, 2016.

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.: Projecting Antarctic ice discharge using response functions from SeaRISE ice-sheet models, Earth System Dynamics, 5, 271–293, 2014.

Little, C. M., Oppenheimer, M., and Urban, N. M.: Upper bounds on twenty-first-century Antarctic ice loss assessed using a probabilistic framework, Nature Climate Change, 3, 654, https://doi.org/10.1038/nclimate1845, 2013.

MacAyeal, D. R., Bindschadler, R. A., and Scambos, T. A.: Basal friction of ice stream E, West Antarctica, Journal of Glaciology, 41, 247–262, https://doi.org/10.3189/S0022143000016154, 1995.

McMillan, M., Shepherd, A., Sundal, A., Briggs, K., Muir, A., Ridout, A., Hogg, A., and Wingham, D.: Increased ice losses from Antarctica detected by CryoSat-2, Geophysical Research Letters, 41, 3899–3905, https://doi.org/10.1002/2014GL060111, 2014.

Naughten, K. A., Meissner, K. J., Galton-Fenzi, B. K., England, M. H., Timmermann, R., and Hellmer, H. H.: Future projections of Antarctic ice shelf melting based on CMIP5 scenarios, Journal of Climate, 31, 5243–5261, 2018.

Nias, I., Cornford, S., and Payne, A.: New mass-conserving bedrock topography for Pine Island Glacier impacts simulated decadal rates of mass loss, Geophysical Research Letters, 45, 3173–3181, https://doi.org/10.1002/2017GL076493, 2018.

Nias, I., Cornford, S., Edwards, T., Gourmelen, N., and Payne, A.: Assessing uncertainty in the dynamical ice response to ocean warming in the Amundsen Sea Embayment, West Antarctica, Geophysical Research Letters, 46, https://doi.org/10.1029/2019GL084941, 2019.

Nias, I. J., Cornford, S. L., and Payne, A. J.: Contrasting the modelled sensitivity of the Amundsen Sea Embayment ice streams, Journal of Glaciology, 62, 552–562, https://doi.org/10.1017/jog.2016.40, 2016.

O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, Reliability Engineering & System Safety, 91, 1290–1300, https://doi.org/10.1016/j.ress.2005.11.025, 2006.

Pattyn, F.: The paradigm shift in Antarctic ice sheet modelling, Nature communications, 9, 2728, https://doi.org/10.1038/s41467-018-05003-z, 2018.

Pattyn, F., Favier, L., Sun, S., and Durand, G.: Progress in numerical modeling of Antarctic ice-sheet dynamics, Current Climate Change Reports, 3, 174–184, https://doi.org/10.1007/s40641-017-0069-7, 2017.

Petra, N., Martin, J., Stadler, G., and Ghattas, O.: A computational framework for infinite-dimensional Bayesian inverse problems, Part II: Stochastic Newton MCMC with application to ice sheet flow inverse problems, SIAM Journal on Scientific Computing, 36, A1525–A1555, https://doi.org/10.1137/130934805, 2014.

Pollard, D., Chang, W., Haran, M., Applegate, P., and DeConto, R.: Large ensemble modeling of the last deglacial retreat of the West Antarctic Ice Sheet: comparison of simple and advanced statistical techniques, Geoscientific Model Development, https://doi.org/10.5194/gmd-9-1697-2016, 2016.

Pritchard, H., Ligtenberg, S. R., Fricker, H. A., Vaughan, D. G., van den Broeke, M. R., and Padman, L.: Antarctic ice-sheet loss driven by basal melting of ice shelves, Nature, 484, 502–505, https://doi.org/10.1038/nature10968, 2012.

Pukelsheim, F.: The Three Sigma Rule, The American Statistician, 48, 88–91, https://doi.org/10.1080/00031305.1994.10476030, 1994.

Rasmussen, C. E. and Williams, C. K.: Gaussian processes for machine learning, vol. 2, MIT Press Cambridge, MA, 2006.

Rignot, E., Mouginot, J., and Scheuchl, B.: Ice flow of the Antarctic ice sheet, Science, 333, 1427–1430, https://doi.org/10.1126/science.1208336, 2011a.

Rignot, E., Velicogna, I., van den Broeke, M. R., Monaghan, A., and Lenaerts, J. T.: Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise, Geophysical Research Letters, 38, 2011b.

Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H., and Scheuchl, B.: Widespread, rapid grounding line retreat of Pine Island, Thwaites, Smith, and Kohler glaciers, West Antarctica, from 1992 to 2011, Geophysical Research Letters, 41, 3502–3509, https://doi.org/10.1002/2014GL060140, 2014.

Ritz, C., Edwards, T. L., Durand, G., Payne, A. J., Peyaud, V., and Hindmarsh, R. C.: Potential sea-level rise from Antarctic ice-sheet instability constrained by observations, Nature, 528, 115, https://doi.org/10.1038/nature16147, 2015.

Ruckert, K. L., Shaffer, G., Pollard, D., Guan, Y., Wong, T. E., Forest, C. E., and Keller, K.: Assessing the impact of retreat mechanisms in a simple Antarctic ice sheet model using Bayesian calibration, PLoS One, 12, e0170 052, https://doi.org/10.1371/journal.pone.0170052, 2017.

Salter, J. M. and Williamson, D. B.: Efficient calibration for high-dimensional computer model output using basis methods, arXiv preprint arXiv:1906.05758, 2019.

Salter, J. M., Williamson, D. B., Scinocca, J., and Kharin, V.: Uncertainty quantification for computer models with spatial output using calibration-optimal bases, Journal of the American Statistical Association, pp. 1–40, 2018.

Schlegel, N.-J., Seroussi, H., Schodlok, M. P., Larour, E. Y., Boening, C., Limonadi, D., Watkins, M. M., Morlighem, M., and Broeke, M. R.: Exploration of Antarctic Ice Sheet 100-year contribution to sea level rise and associated model uncertainties using the ISSM framework, The Cryosphere, 12, 3511–3534, https://doi.org/10.5194/tc-12-3511-2018, 2018.

Schoof, C. and Hindmarsh, R. C.: Thin-film flows with wall slip: an asymptotic analysis of higher order glacier flow models, Quarterly journal of mechanics and applied mathematics, 63, 73–114, 2010.

Sexton, D. M., Murphy, J. M., Collins, M., and Webb, M. J.: Multivariate probabilistic projections using imperfect climate models part I: outline of methodology, Climate dynamics, 38, 2513–2542, https://doi.org/10.1007/s00382-011-1208-9, 2012.

Shepherd, A., Ivins, E., Rignot, E., Smith, B., Van Den Broeke, M., Velicogna, I., Whitehouse, P., Briggs, K., Joughin, I., Krinner, G., et al.: Mass balance of the Antarctic Ice Sheet from 1992 to 2017, Nature, 558, 219–222, https://doi.org/10.1038/s41586-018-0179-y, 2018.

Vernon, I., Goldstein, M., Bower, R. G., et al.: Galaxy formation: a Bayesian uncertainty analysis, Bayesian analysis, 5, 619–669, https://doi.org/10.1214/10-BA524, 2010.

Williamson, D. B., Blaker, A. T., and Sinha, B.: Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model, Geoscientific Model Development, 10, 1789–1816, https://doi.org/10.5194/gmd-10-1789-2017, 2017.

Yu, H., Rignot, E., Seroussi, H., and Morlighem, M.: 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, 3861–3876, https://doi.org/10.5194/tc-12-3861-2018, 2018.