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 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 dimension-reduced approach to calibration combined with statistical emulation of the adaptive mesh model BISICLES. We find the most likely contribution to global mean sea level rise from the Amundsen Sea Embayment (ASE) over the next 50 years is 10.4 [0.6, 23.3] mm (mode and 5-95% probability interval), a substantial reduction in uncertainty from the uncalibrated estimates of 9.6 [-5.9, 78.2] mm. We predict retreat of the grounding line along most parts of the ASE coast with high confidence, with a maximum inland extent of around 28 km at Smith Glacier. The 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, though this does influence the future mass loss less than basal traction and viscosity scaling parameters. 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 Antarctic ice sheet on continental and centennial scales by efficiently improving our understanding of model behaviour, and 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/year sea level equivalent, 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

5   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

10 global mean sea level (Fretwell et al., 2013). However, the future response of the Antarctic ice sheet 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

15 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,

20 2016; Schlegel et al., 2018; Gladstone et al., 2012; Ritz et al., 2015; Bulthuis et al., 2019) (Here we do not distinguish between initial values of state variables, which will change during the simulation, and 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, necessitating parameterisation of grounding line retreat, or else restricts the domain to a single glacier. One exception is the

25 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 by the end of the century typically ranges from around zero to tens of centimetres, i.e. the ensemble spread is twice the predicted (mean/median) contribution (Edwards et al., 2019). It is therefore essential to constrain ice sheet model parameters to reduce these uncertainties i.e. to attain sharper

30 and more distinctive prediction distributions for different climate scenarios. 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 treat-
5 ments (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
10 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, often into a single value (Ritz et al., 2015; DeConto and Pollard, 2016; Edwards et al., 2019), a more
15 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 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
20 formal probabilistic approach, still assume spacial 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
25 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, essentially by interpolation in the parameter space. This approach has only recently been applied in projections of the Antarctic ice sheet contribution to sea level rise (Edwards et al., 2019; Chang et al., 2016a, b). Emulation becomes particularly important in model calibration, as this down-weights or rejects ensemble members and therefore reduces the effective ensemble size.
30 The aim of this study is to use a novel, practical, yet comprehensive calibration of the high-resolution Antarctic ice sheet model BISICLES to give smooth, refined probability functions for the dynamic sea level contribution from 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 inference for probabilistic predictions. We use emulation of
35 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 in Section 3. In Section 4 we present the resulting probabilistic ice sheet predictions which are discussed in Section 5.

2 Model Ensemble and Observations

2.1 Ice sheet model ensemble

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. 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 of the model by Nias et al. (2016), using inversion from surface ice speeds (Rignot et al., 2011), and subsequently perturbed between half and double the default values in a Latin Hypercube design (Nias et al., 2016). We use normalized parameter ranges with halved, default and doubled scaling factors mapped to 0, 0.5 and 1, respectively. In addition, two further uncertain inputs are varied categorically: two different bedrock elevation maps are used, as well as two different sliding 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 law exponent defines the linearity of the basal ice velocity with basal traction, and values of 1 (linear) and 1/3 (power law) are used. For a full description of the model ensemble see Nias et al. (2016). For the calibration we regrid the simulated surface elevation fields for the first 7 years to the same spatial resolution as the observations (10 km×10 km) by averaging.

We do use the whole model domain and a finer model resolution of 4 km×4 km for projections (after 50 years) because we are not restricted by the observational data characteristics. Besides this, the same methods (spatial decomposition and emulation, see below) are used for the projections (after 50 years) as for the calibration period. The spatial mask defining the catchment area as in Nias et al. (2016) is used for sea level rise estimates.

2.2 Observations

We use 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.

5   3   Theoretical basis and Calibration Model

Our calibration approach consists of an emulation step 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 emulation and calibration take place in the basis representation of a Principal Component (PC) decomposition, in order to adequately

10   represent spatial correlation and avoid unnecessary loss of information (e.g. by comparing total or mean model-observation differences). We build two spatial emulators: one represents the model response after the first 7 years of the simulation, which is used for calibration. The second emulator represents the response at the end of the 50 year simulation, and is used to make probabilistic predictions of total sea level contribution and spatial grounding line retreat.

3.1 Principal Component Decomposition

15   Let $\boldsymbol{y}(\boldsymbol{\theta_i})$ be the m dimensional spatial ice sheet model output for a parameter setting $\boldsymbol{\theta_i}$, where m is the number of spatial grid cells and the model ensemble has n members so that $\boldsymbol{\theta_1}, ..., \boldsymbol{\theta_n} \in \boldsymbol{\Theta}$, $\boldsymbol{\Theta} \in R^d$ being the parameter space which is in our case spanned by $d = 5$ model input parameters. The $m \times n$ matrix $\widetilde{\mathbf{Y}}$ is the row-centered combined model output 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. A principal component decomposition is achieved by finding $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$ so that

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

where the $n \times n$ 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 eigenvectors of $\widetilde{\mathbf{Y}}^T\widetilde{\mathbf{Y}}$ and the columns of $\mathbf{U}$ are the 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 and we call the i.th column of $\mathbf{U}$ the i.th principal component.

25   The fraction of ensemble variance represented by a principal component is proportional to its eigenvalue and typically there is a number $k < n$ for which the first k principal components represent the whole ensemble sufficiently well. We choose $k = 4$ for the emulator at the beginning of the model period and k=5 for the second emulator at the end of the period so that in both cases 90% of the model variance is captured (Appendix A).

$$\widetilde{\mathbf{Y}} \approx \mathbf{U}'\mathbf{S}'\mathbf{V}'^T \tag{2}$$
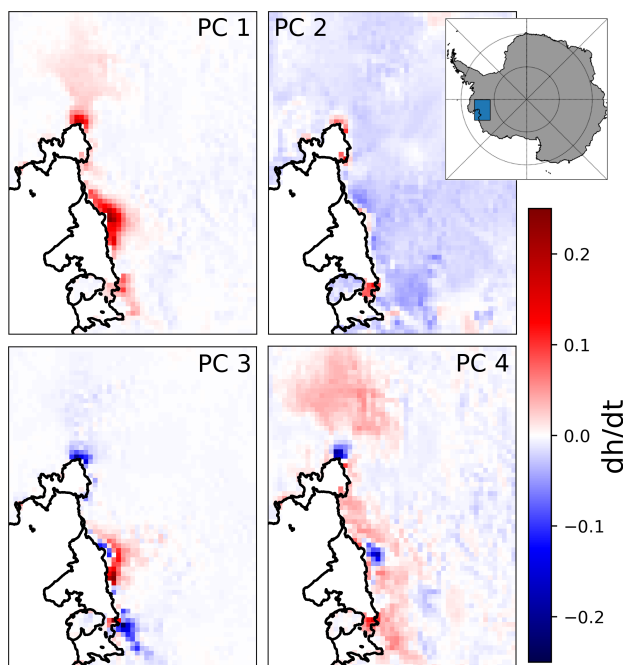
Figure 1. First four PCs of the calibration period with unit length

with $\mathbf{U}'$ and $\mathbf{V}'$ consisting of the first k columns of $\mathbf{U}$ and $\mathbf{V}$, while $\mathbf{S}'$ is the $k \times k$ upper left corner matrix of $\mathbf{S}$

This decomposition reduces the dimensions from m grid cells to just k principal components. The PCs are by construction orthogonal to each other and can be treated as statistically independent.

3.2 Observations in basis representation

5 Spatial m dimensional observations $\boldsymbol{z}_{(\boldsymbol{xy})}$ can be transformed to the basis representation by:

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

for $\boldsymbol{z}_{(\boldsymbol{xy})}$ 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 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

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

Figure 2 shows that large parts of the observations can be represented by the first 4 PCs from Fig. 1. It is only this part illustrated on the right of Fig. 2 which is used for calibration. The spatial variance of the difference between the reprojected and original fields is substantially smaller than from $\boldsymbol{z}_{(\boldsymbol{xy})}$ alone:
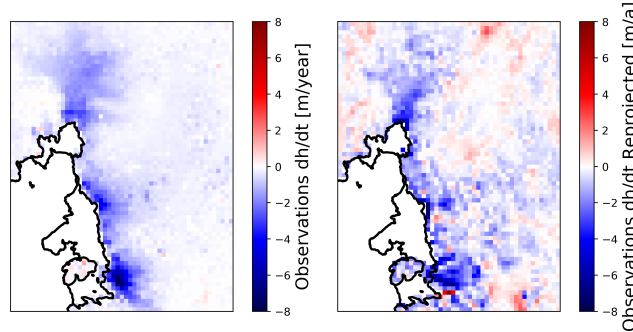
Figure 2. Left: Mean observations. Right: Observations projected to first 4 PCs and reprojected to spatial field

$$\frac{VAR(\boldsymbol{z_{(xy)}} - \mathbf{U}'((\mathbf{U}'^T\mathbf{U}')^{-1}\mathbf{U}'^T\boldsymbol{z_{(xy)}}))}{VAR(\boldsymbol{z_{(xy)}})} \approx 0.62$$

3.3 Emulation

For probabilistic prediction 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

5 input space. In our case more than half of the projection distribution would be based on just six BISICLES ensemble members. For computationally expensive models sufficient sampling can be achieved by statistical emulation, as laid out in the following.

We use each row of $\mathbf{S}'\mathbf{V}'^T$ in combination with $\boldsymbol{\Theta}$ to train independent continuous statistical models. Each of those statistical 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 call the aggregation of those statistical models an emulator. We use

10 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). We refer to Rasmussen and Williams (2006) for an in-depth discussion and tutorial of Gaussian Process Emulators. The emulator can be described as a joint Gaussian distribution:

$$\Omega = N(\boldsymbol{\omega}(\boldsymbol{\theta}), \quad \boldsymbol{\Sigma}_\omega(\boldsymbol{\theta})) \tag{4}$$

15 where $\boldsymbol{\Sigma}_\omega$ is the $k \times k$ emulator co-variance matrix and $\boldsymbol{\omega}$ the emulator mean vector for any given $\boldsymbol{\theta}$. Here $\boldsymbol{\Sigma}_\omega$ is diagonal due to the statistical independence of the principal components. The diagonal values of $\boldsymbol{\Sigma}_\omega$ are derived from the corresponding Gaussian Process models and are dependant on the distance between the new $\boldsymbol{\theta}$, for which the emulator is evaluated, and the training points from the ice sheet model. The nature of this distance dependency is defined by the Gaussian Process covariance function which we define as Matern ($\frac{5}{2}$) type with length scales optimized on the marginal likelihood. In total we generate more

20 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 $R^2 > 0.993$ (Appendix A).

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 (Chang et al., 2016a; Salter et al., 2018, e.g.) or in the basis representation itself (Higdon et al., 2008, e.g.). Here we will base probability statements on the PC basis representation.

We assume the existence of a parameter configuration $\boldsymbol{\theta}^*$ within the bounds of $\boldsymbol{\Theta}$ 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:

$$\pi(\boldsymbol{\theta}|\boldsymbol{z}=\hat{\boldsymbol{z}}) \propto L(\boldsymbol{z}=\hat{\boldsymbol{z}}|\boldsymbol{\theta}) \times \pi(\boldsymbol{\theta}) \tag{5}$$

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)). We relate the observables to the real state of the ice sheets in basis representation, $\boldsymbol{\gamma}$, by:

$$\boldsymbol{z} = \boldsymbol{\gamma} + (\mathbf{U}'^T\mathbf{U}')^{-1}\mathbf{U}'^T\boldsymbol{e} \tag{6}$$

where $\boldsymbol{e}$ is the spatial observational error and the transformation $(\mathbf{U}'^T\mathbf{U}')^{-1}\mathbf{U}'^T$ follows from equ 3. The emulator output is related to $\boldsymbol{\gamma}$ by the model discrepancy $\boldsymbol{\varepsilon}$:

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

We simplify the probabilistic inference by assuming the observational error $\boldsymbol{e}$, model error/discrepancy $\boldsymbol{\varepsilon}$ and the model parameter values $\boldsymbol{\Omega}$ to be mutually statistically independent with $\boldsymbol{\varepsilon} = N(0, \quad \boldsymbol{\Sigma}_\varepsilon)$. We further assume the observational error to be spatially identically distributed with variance σ_e^2, so that

$$(\mathbf{U}'^T\mathbf{U}')^{-1}\mathbf{U}'^T\boldsymbol{e} = N(0, \quad \sigma_e^2(\mathbf{U}'^T\mathbf{U}')^{-1}) \tag{8}$$

The $k \times k$ matrix $(\mathbf{U}'^T\mathbf{U}')^{-1}$ is diagonal with the element-wise inverse of $diag(\mathbf{S}')_i^2$ as diagonal values. We estimate σ_e^2 from the variance among the 14 observational periods for the first principal component constituting $\hat{z}_1$, i.e.

$$\sigma_e^2 = VAR(\hat{z}_1) \cdot diag(\mathbf{S}')_1^2 \tag{9}$$

Note that the existence of $\boldsymbol{\gamma}$ is an abstract concept, implying that it is only because of an error $\boldsymbol{\varepsilon}$ 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 uncertainty, whether explicitly by setting $\boldsymbol{\varepsilon} = \mathbf{0}$, or implicitly, would imply that the 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.

It follows from equations 4, 6, 7 and 8 that

$$5 \quad L(\boldsymbol{z} = \hat{\boldsymbol{z}}|\boldsymbol{\theta}) \propto exp\left[-\frac{1}{2}(\boldsymbol{\omega}(\boldsymbol{\theta}) - \hat{\boldsymbol{z}})^T \boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\omega}(\boldsymbol{\theta}) - \hat{\boldsymbol{z}}) \right] \tag{10}$$

with $\boldsymbol{\Sigma}_T = \sigma_e^2 (\mathbf{U'}^T \mathbf{U'})^{-1} + \boldsymbol{\Sigma}_\varepsilon + \boldsymbol{\Sigma}_\omega$.

The inclusion of systematic model uncertainties can lead to identifiability issues where the model signal cannot be distinguished from imposed systematic model uncertainty (discrepancy). To overcome such issues, constraints on the, e.g. spatial shape, of the discrepancy can be used (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 $diag(\boldsymbol{\Sigma}_\varepsilon) = (\sigma_{\varepsilon 1}^2, ..., \sigma_{\varepsilon k}^2)^T$. A common rule of thumb is that at least 95% of a probability distribution lies within three standard deviations from the mean (Pukelsheim, 1994). For the i.th PC we find $\sigma_{\varepsilon i}^2$ so that 95% of the observational distribution $N(\hat{z}_i, \quad \sigma_{ei}^2)$ lies within $3\sigma_{\varepsilon i}$ from the mean of $\boldsymbol{\omega}(\boldsymbol{\Theta})_i$, i.e. across the n ensemble members. We thereby force the observations to fulfill the 'three-sigma rule' by considering them as part of the model distribution $\boldsymbol{\omega}(\boldsymbol{\Theta})_i$.

The calibration distribution in Equation 5 can be evaluated using Eq. 10 with a trained emulator (Eq. 4), observational (Eq. 9) and model uncertainties (above) and the prior parameter distributions $\pi(\theta)$ set by expert judgment.

4 Results

4.1 Parameter calibration

Figure 3 illustrates which parts of the model input space are most successful in reproducing the satellite observations of surface elevation changes during the initial part of the simulation. 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. 3 it is not possible to see 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.

The calibration has the strongest effect on selecting the bedrock topography map and sliding law exponent. We find that the modified bedrock from Nias et al. (2016) produces much more realistic surface elevation changes than the original Bedmap2
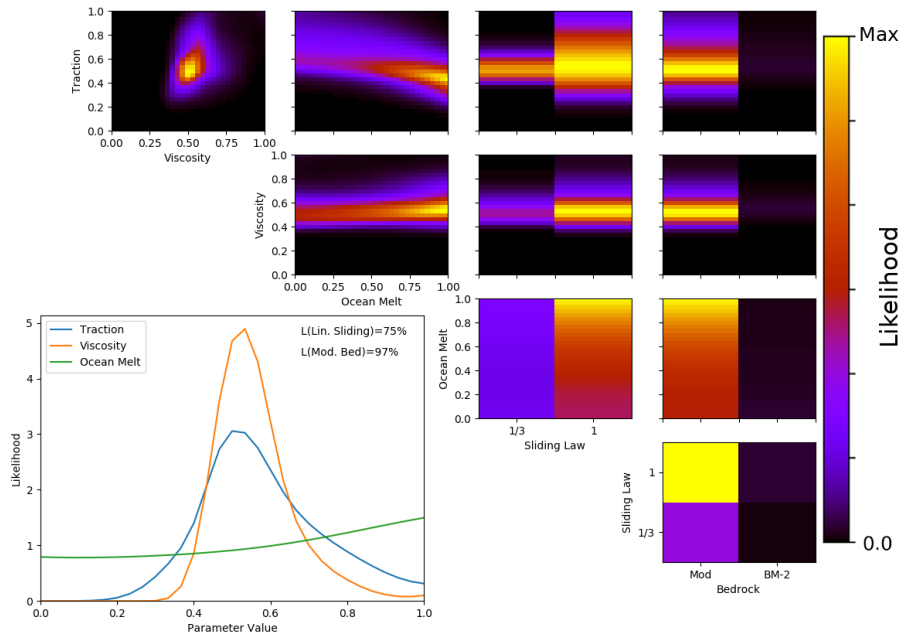
Figure 3. Likelihood of different model inputs (evaluations of Equation 10). Upper right panels show likelihood values marginalized to pairs of parameters, normalized to the respective maximum for clarity. Lower left panel shows likelihood values marginalized to individual parameters for the three scalar parameters (line plots), and sliding law and bedrock topography map (text and quotation within), normalized to an integral of one in the style of Probability Density Functions.

topography (Fig. 3), and the linear sliding law is more successful than the non-linear ($m = \frac{1}{3}$). We find the most likely fields of basal traction and velocity are the default values (i.e. scaling factor of 0.5) inferred by (Nias et al., 2016) from inversion from surface ice speeds; this confirmation through two independent datasets suggests good model consistency. The sub-shelf melting field is constrained the least (all values of the scaling factor are similarly likely), probably due to the short calibration

5 period, but higher melt rates are found to be slightly more likely than lower rates. Nias et al. (2016) find the model sensitivity to ocean melt to be considerably smaller than to viscosity and traction, in particular for the large Thwaites Glacier. The parameter combination with the highest likelihood has values of 0.43 for basal traction, 0.53 for viscosity, and 1 (maximum) for ocean melt factor, with a linear sliding law and the modified bedrock topography. However, for probabilistic predictions we do not focus on just this one parameter configuration but use the whole input space weighted by probability.

10 **4.2 Sea Level Contribution Projection**

We use the calibration shown in Fig. 3 to update the predictions of sea level contribution and grounding line retreat after 50 years.
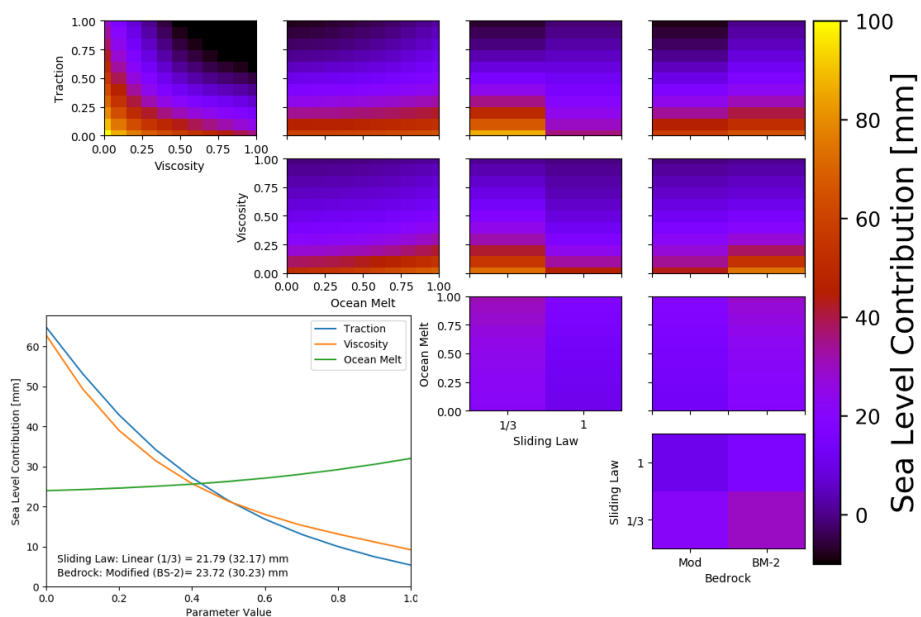
Figure 4. Mean sea level contributions at end of 50 year model period as in Fig. 3. For each parameter combination (upper right pannels) or single parameter value (lines) the mean sea level contributions accross all other parameters is calculated using the prior dirstibutions (i.e. equal weights)
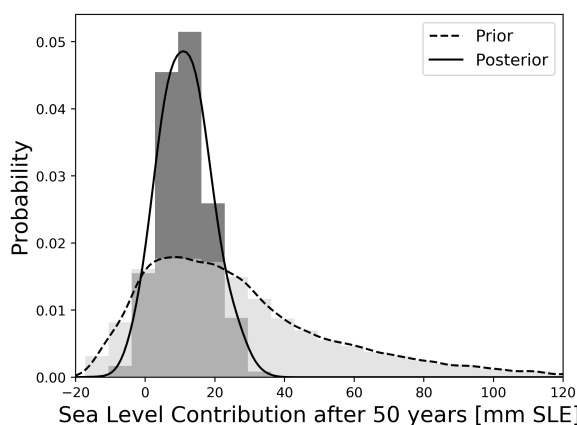
Figure 5. Total sea level contribution from the Amundsen Sea Embayment after 50 years. The full (shaded) and probability weighted (intense colors) emulated ensemble distributions are shown.

The calibrated prediction distribution is centred around the mode of the uncalibrated distribution, but sharper (i.e. reduced uncertainty) and much more symmetric (Fig. 5 and Table 1). This is most striking at the upper end of the distribution: the 95th percentile decreases from 78.2 to 23.3 mm SLE. We can also essentially rule out (<5%) negative sea level contributions.

Table 1. Total sea level contribution in mm SLE: (weighted) mean, most likely model configuration and percentiles; with and without calibration.

	Mean	Mode	$\max(L(\boldsymbol{z} = \hat{\boldsymbol{z}}|\boldsymbol{\theta}))$	5%	25%	50%	75%	95%
Prior year 50	25.6	9.6	—	-5.9	6.8	20.2	37.9	78.2
Posterior year 50	11.4	10.4	19.2	0.6	6.3	11	16.1	23.3

The tail is mostly rejected by downweighting of model simulations in which both the basal traction and basal viscosity scaling factors are small, because these give unrealistically large surface elevation decreases (compare Fig. 3 and 4). The second most important factor is the preference for the linear sliding law; again, the power law tends to give greater ice thinning rates than linear sliding. Bedrock topography is less important than the sliding law but the use of the modified bedrock in the uncalibrated

5   distribution reduced the mean sea level contribution nevertheless by about 20% (Fig. 4; Nias et al., 2016).

The sea level contribution from the best (maximum likelihood) ensemble member over the 50 year period is 19.2 mm, notably higher than the distribution mean, mode and median (10-11 mm SLE). This ensemble member has the highest possible value for the ocean melt factor (1). This causes more ensemble members to have smaller rather than larger sea level contributions hence influencing the mean and median in this direction.

10   **4.3 Grounding line retreat probabilities**

Figure 6 shows the probabilities of regions to become ungrounded for the prior (uncalibrated) and posterior (calibrated) distributions. The area potentially affected by a grounding line retreat is considerably smaller after calibration, which is consistent with the reduced upper tail of total sea level contribution in Fig. 5. In other words, substantial grounding line retreat is simulated with low basal traction and viscosity, or a power law basal sliding law, but these are found to be much less consistent with

15   recent surface elevation observations. However, all glaciers (Pine Island, Thwaites and Smith/Pope) retreat to some degree, and the farthest retreat is ≈ 28 km for Smith Glacier.

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.,

20   2018; Bulthuis et al., 2019; Cornford et al., 2015), 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) model predictions of both sea level contribution and grounding-line retreat.

25   Our Bayesian calibration of the sliding law with surface elevation changes is consistent with the Bayesian calibration with ASE mass loss of (Ritz et al., 2015), who find that the linear sliding law is more likely to produce mass loss rates consistent
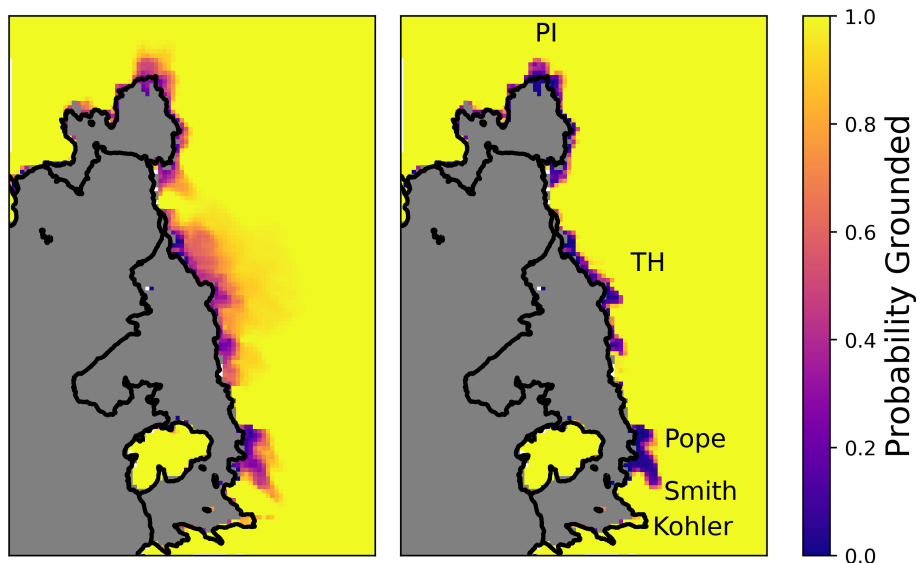
Figure 6. Probability density estimates of grounding line retreat after 50 years, uncalibrated (left) and calibrated (right). Labeled glaciers include Pine Island (PI) and Thwaites (TH). Initially ungrounded areas are masked in gray to highlight retreat, black lines show estimates of the grounding line and ice cliffs (MODIS Mosaic of Antarctica 2009 (Scambos et al., 2007))

.

with observed than nonlinear and plastic laws. However, Joughin et al. (2009) find a linear law underestimates sensitivity to basal traction.

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 predicted sea level contribution from Pine Island Glacier for some sliding laws (Nias et al., 2018). Here we find the modified bedrock topography far more consistent with observed surface elevation changes than the original Bedmap2 bedrock (Fig. 3) while a calibration with total mass loss alone cannot distinguish between the two (see Figure 2 of Nias et al. (2016)). Calibrated predictions for this region based on Bedmap2 are likely to either be compensating the overly-stabilising bedrock with underestimated viscosity and/or traction coefficients, or underestimating the sea level contribution altogether. Note that the modified bedrock has been derived by reducing clearly unrealistic behaviour of the same ice sheet model, a better calibration performance is therefore to be expected. However, no satellite observations have been used for the bedrock modification, nor has there a been a quantitative probabilistic assessment.

Satellite-based estimates for the Amundsen Sea Embayment of 0.33 mm SLE per year (McMillan et al., 2014) from 2010-2013 are higher than the mode of our sea level contribution distribution of about 0.21 mm SLE per year but similar to the 75th percentile.

Even though the most extreme retreat scenarios are essentially ruled out by the calibration, we see several locations with high probability to become ungrounded. Most prominent is the retreat of Smith Glacier (Fig. 6). Previous studies have shown how

far the grounding line could retreat for different model ensembles and climate scenarios (Nias et al., 2016; Bulthuis et al., 2019; Ritz et al., 2015), but only (Ritz et al., 2015) provide calibrated probabilistic estimates to compare with here. The resolution of their model is far lower (15 km), but the pattern approximately corresponds to their 95% probability contour by 2100 for Pine Island and Thwaites Glacier. Comparison with present day observations of grounding line retreat for Thwaites Glacier

5 (for 2018; Milillo et al., 2019) and the Pope/Smith/Kohler region (for 2014; Scheuchl et al., 2016) show good agreement of the locations of retreat (e.g. far retreat at Smith with minimal retreat of Kohler glacier). However, large parts of the retreat predicted in this study for the 50 year period (starting somewhere between year 2000 and 2010) has been reached already. There are two possible explanations for this: the grounding line has reached a new stable location supported by topographic features (which are visible in Milillo et al. (2019) and Scheuchl et al. (2016)), where it will remain for the next decades, or else

10 we underestimate the rate of retreat and with it potentially the rate of mass loss.

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. Warmer ocean and air temperatures would enhance melt and accelerate the dynamic response. Neither do the simulations carry the countervailing predicted increase of surface accumulation in a warmer climate (Lenaerts et al., 2016). Edwards et al. (2019); Golledge et al. (2019) found that the Antarctic ice sheet response to very different

15 greenhouse gas emissions scenarios starts to diverge from around 2060-2070, indicating that scenario would have a small net impact on our 50-year projections.

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 (PC) representation has been adapted and tested for general circulation (climate) and ocean models Sexton et al. (2012); Chang et al. (2014); Salter

20 et al. (2018); Salter and Williamson (2019). 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. Our method is novel because we calibrate a grounding line resolving ice sheet model in the PC space, to avoid the assumption that the difference between observation and calibration model are spatially uncorrelated (e.g. Chang et al., 2016b)). In comparison with studies that use highly aggregated quantities (like total sea level contribution (e.g. Ritz et al., 2015)), we are able to

25 exploit more of the available observational information to add further constraints to the input parameters and sharpen the posterior distribution. Appendix B shows a calibration with total sea level contribution: the parameters are less constrained, particularly the bedrock topography and the allowable combinations of basal traction and viscosity.

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 Fig 2). This can mean that there is no parameter configuration θ which

30 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 (Fig. 2b) 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.

6   Conclusions

We present probabilistic estimates of the dynamic contribution to sea level of the Amundsen Sea Embayment in West Antarctica over the next 50 years using a grounding line resolving ice sheet model. We performed a Bayesian calibration of a published ice sheet model ensemble with satellite measurements of surface elevation changes from 1992-2015, using spatial decomposition

5   to increase the amount of information used from the observations and emulation techniques to search the parameter space more thoroughly and estimate the probability distribution for sea level.

We find that the modified bedrock topography derived by Nias et al. (2016) results in quantitatively far more consistent model representation of the Amundsen Sea Embayment than Bedmap2. Identifying the most successful basal sliding law and ocean melt rate is more challenging probably due to their slow impact on ice sheet behaviour and the short calibration timescale.

10   Nevertheless, we find the calibration favours higher melt rates than those derived by Nias et al. (2016) in model initialisation and a linear sliding law.

The calibration leads to a substantial reduction in the upper tail of the probability distribution of sea level contribution, leading to quite a symmetric distribution. The predicted sea level contribution within the next 50 years from the Amundsen Sea Embayment will most likely be between 0.6 and 23.3 mm SLE (90% probability interval) with the median and most likely

15   model parameter configurations predicting 11.0 and 19.2 mm SLE, respectively. The calibration limits the predicted extent of grounding line retreat for Pine Island and Thwaites Glaciers, but predicts up to 28 km retreat for Smith Glacier.

Code availability.   On Request

Appendix A: Emulator regression and validation

We use Gaussian Process (GP) models for emulation and train a separate GP model for each Principal Component (PC) and

20   time period. We use a Matern $(\frac{5}{2})$ covariance function, with the covariance function (hyper-) parameters being optimized on the marginal likelihood with five repetitions using the Python GPy module. The nugget term is set to zero, forcing the emulator to predict the exact values at training points, reflecting the deterministic nature of the ice sheet model. A constant mean function with $N(0, 0.5)$ prior is used, accounting for the initial centering of $\widetilde{\mathbf{Y}}$.

The first 4 PCs are emulated for the calibration (first 7 years of model period) and 5 PCs are emulated for the predictions

25   (after 50 years). This choice is based on the different decline in variance represented by PCs and ensures that 90% of the variance is covered in both cases (Fig. A1).

In the following we will illustrate the emulator performance by a leave-one-out (LOO) cross-validation scheme. For this we repeat all steps of the the emulator setup for subsets of all but one of the full ensemble, and use that emulator to predict the PC scores of the left-out ensemble member. These are compared with the actual ice sheet model values to validate the emulator.

30   This process is repeated until each ensemble member is left out once.
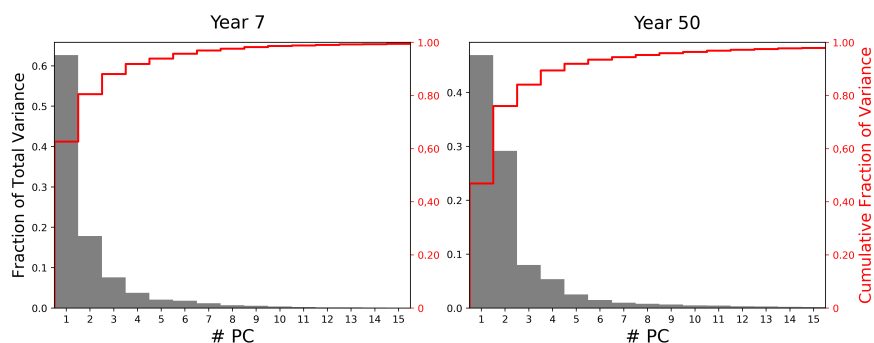
Figure A1. Variance in model ensemble ice thickness change fields explained by PCs after 7 (left) and 50 years (right). The illustrated variance scales with the eigenvalues of corresponding PCs which are the squared singular values (diagonal entities of **S**).
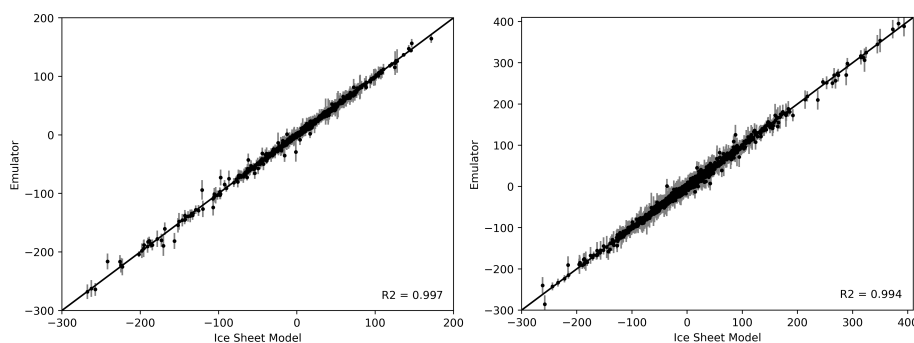
Figure A2. Leave-one-out emulator validation plot for year 7 (left) and year 50 (right). Grey error bars represent $3\,\sigma_\omega$, i.e. emulator uncertainties. All $k = 4$ (left) and $k = 5$ (right) PC scores of each LOO repetition are shown together.

Figure A2 shows the ice sheet model PC scores versus the LOO emulator prediction of the same quantity. We see an overall good correlation without serious outliers. The emulator uncertainty is assessed as well in Table A1. Around 90% of the differences between emulator and ice sheet model are within the two σ_ω emulator uncertainty intervals, i.e. approximately as expected (95%) for a normal distribution. The emulator performance is uniform within the parameter space (not shown).

5 **Appendix B: Net sea level contribution calibration**

Figure B1 shows the result of calibrating with total sea level rise contribution of 0.33 mm SLE per year (McMillan et al., 2014), instead of two-dimensional surface elevation changes in the PC space as in the main analysis (Fig. 3). The parameter space is less constrained, particularly for bedrock topography, where neither choice is strongly preferred, and the combination of basal traction and viscosity, where more combinations are allowed. However, the two calibrations do not contradict each other, as 10 there is a considerable amount of intersection in the estimates of likely parameter combinations. This demonstrates the value of the extra information - and stronger parameter constraints - provided by the use of two-dimensional observations.

Table A1. Emulator validation metrics

	Year 7	Year 50
RMSE (predicted-simulated)	3.07	4.76
RMSE (predicted-simulated)/range	0.70%	0.73%
Pearson's r	0.998	0.997
Spearman's rho	0.998	0.993
Kendall tau	0.970	0.940
Fraction in 95% range	89.1%	90.5%
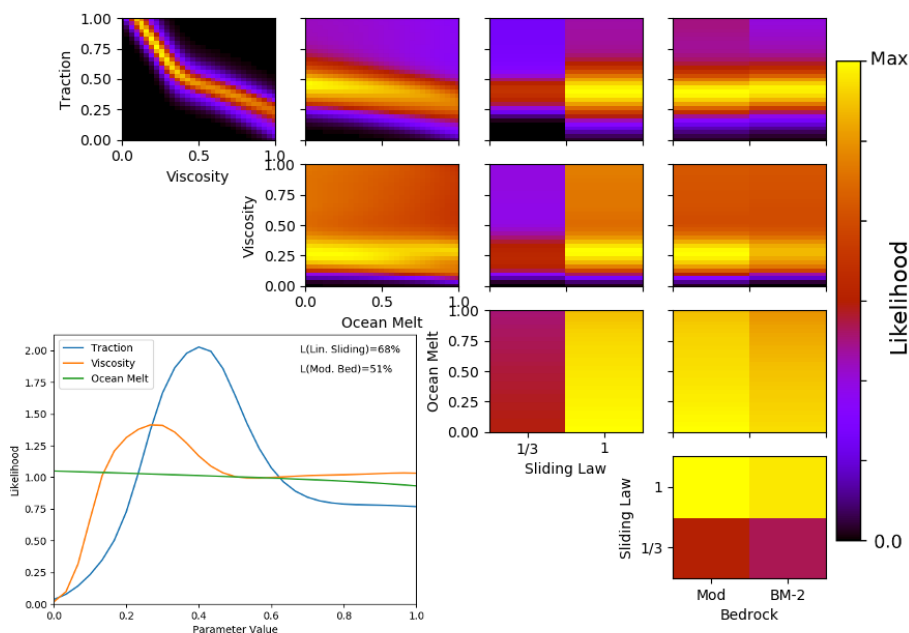
Figure B1. Likelihood evaluations as in Fig. 3 but using exclusively the seven-year mean rate of sea level rise contribution from the observed area for likelihood calculations. The observational uncertainty is based on the variance among the 14 observational periods and the model discrepancy σ_ε^2 is set to twice the observational variance while emulator uncertainty is neglected here

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 shareing and advising on the satellite observations and Mark Brandon for general advice.

References

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, 2018.

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, 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, 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, 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, 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, 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, 2015.

DeConto, R. M. and Pollard, D.: Contribution of Antarctica to past and future sea-level rise, Nature, 531, 591, 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, 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.

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, 2013.

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, 2012.

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, 2019.

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, 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, 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, 2015.

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, 2009.

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, 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, 2016.

5 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, 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, 2016.

Levermann, A., Winkelmann, R., Nowicki, S., Fastook, J. L., Frieler, K., Greve, R., Hellmer, H. H., Martin, M. A., Meinshausen, M.,
10 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, 2013.

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

Milillo, P., Rignot, E., Rizzoli, P., Scheuchl, B., Mouginot, J., Bueso-Bello, J., and Prats-Iraola, P.: Heterogeneous retreat and ice melt of Thwaites Glacier, West Antarctica, Science advances, 5, eaau3433, 2019.

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.

20 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, 2018.

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, 2016.

O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, Reliability Engineering & System Safety, 91, 1290–1300, 2006.

25 Pattyn, F.: The paradigm shift in Antarctic ice sheet modelling, Nature communications, 9, 2728, 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, 2017.

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, 2016.

30 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., 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, 2011.

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

5 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.

Scambos, T., Haran, T., Fahnestock, M., Painter, T., and Bohlander, J.: MODIS-based Mosaic of Antarctica (MOA) data sets: Continent-wide surface morphology and snow grain size, Remote Sensing of Environment, 111, 242–257, 2007.

Scheuchl, B., Mouginot, J., Rignot, E., Morlighem, M., and Khazendar, A.: Grounding line retreat of Pope, Smith, and Kohler Glaciers, West
10 Antarctica, measured with Sentinel-1a radar interferometry data, Geophysical Research Letters, 43, 8572–8579, 2016.

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, 2018.

Schoof, C. and Hindmarsh, R. C.: Thin-film flows with wall slip: an asymptotic analysis of higher order glacier flow models, Quarterly
15 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, 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, 2018.

20 Vernon, I., Goldstein, M., Bower, R. G., et al.: Galaxy formation: a Bayesian uncertainty analysis, Bayesian analysis, 5, 619–669, 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, 2017.