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. With the help of a published perturbed parameter ice sheet model ensemble of the Amundsen Sea Embayment (ASE), we show how the use of 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 net sea level contribution calibration imposes only weaker constraints.

Projections of the 50 year sea level rise contribution from the current state of the ASE can be narrowed down with satellite observations of recent ice thickness change to 18.9 [13.9, 24.8] mm (median and 90% range) with the proposed spatial calibration approach, compared to 16.8 [7.7, 25.6] for the net sea level calibration and 23.1 [-8.4, 94.5] 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 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 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 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 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, causing challenges in representing the grounding line, or else are restricted regional applications. 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 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 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 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 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, 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 (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 in Section 3. In Section 4 we present the resulting probabilistic ice sheet projections 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. Two 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) 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., 2011), and subsequently perturbed between half and double the default values in a Latin Hypercube design by (Nias et al., 2016). Different default basal traction coefficient fields have been found for each combination of bed topography and sliding law while the default viscosity field only differs between bed geometries (but not sliding laws). We use the normalized parameter ranges with halved, default and doubled scaling factors mapped to 0, 0.5 and 1, respectively.

The ensemble covers a wide range of sea level rise contributions for the 50 year period with the most extreme members reaching -0.19 mm/year and 1.62 mm/year, respectively. About 10% of the ensemble shows an increasing volume above flotation (negative sea level contribution) and the central runs (0.5 for traction, viscosity and ocean melt parameters) contribute 0.27 mm/year (linear sliding) and 0.26 mm/year (nonlinear sliding). The average contributions are generally reasonably close to satellite observations (0.33 ± 0.05 mm/year from 2010-2013 (McMillan et al., 2014)) with 0.30 mm/year for linear sliding and modified bedrock, 0.37 mm/year for linear sliding and Bedmap-2, 0.38 mm/year for nonlinear sliding and modified bedrock and 0.51 mm/year for nonlinear sliding and Bedmap-2 (Nias et al., 2016).

For a full description of the model ensemble see Nias et al. (2016). We allow for a short spin up phase of 3 years for the model to adjust to the perturbations and use the following 7 years as calibration period. Other calibration periods have been tested and show small impact on the results for calibrations in basis representation. We regrid the simulated surface elevation fields for this period to the same spatial resolution as the observations (10 km$\times$10 km) by averaging. The sea level rise contribution at the end of the model period (50 years) is calculated directly on the model grid, using the same catchment area mask as in Nias et al. (2016).

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.

3 Theoretical basis and Calibration Model

In the following we propose a new ice sheet model calibration approach which 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 emulation and calibration take place in the basis representation of a Principal Component (PC) decomposition, in order to adequately represent spatial correlation and avoid unnecessary loss of information (e.g. by comparing total or mean model-observation differences). We construct a spatial emulator for the calibration period to represents the two dimensional model response in ice thickness change. A second, non-spatial emulator represents 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 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

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

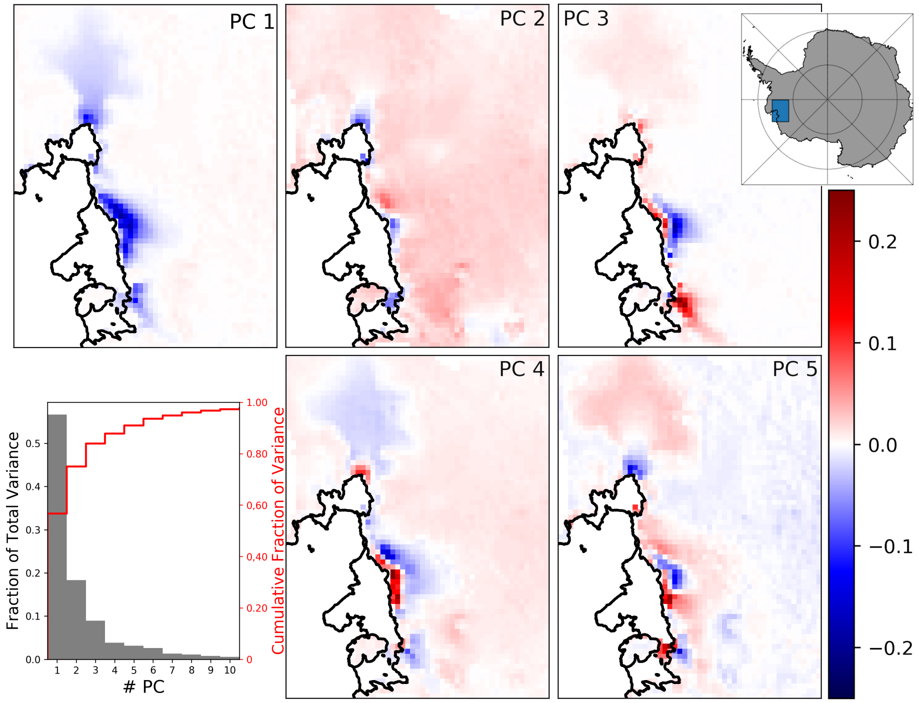

Figure 1. The first five normalized PCs, building an orthogonal basis. They represent the main modes of variation in the model ensemble

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 $\mathbf{B} = \mathbf{US}$ as shorthand for the new basis and call the i.th column of $\mathbf{B}$ the i.th principal component.

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 1). The first k columns of $\mathbf{U}$) are illustrated in Figure 1 which are related to the PCs ($\mathbf{B}_i$) by multiplication with the singular values.

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

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. The PCs are by construction orthogonal to each other and can be treated as statistically independent.

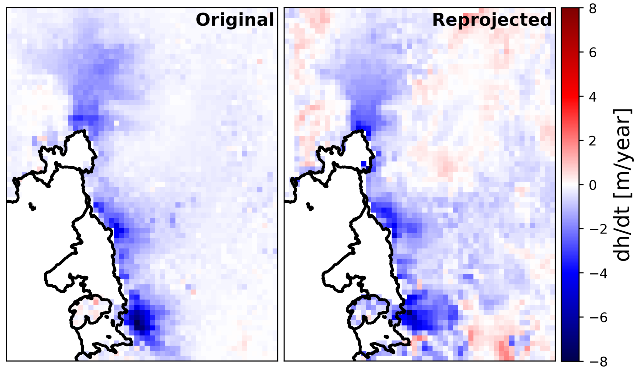

Figure 2. Left: Mean observed ice thickness change. Right: Observed ice thickness change projected to first five PCs and reprojected to spatial field

3.2 Observations in basis representation

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

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

for $\boldsymbol{z_{(xy)}}$ on the same spatial grid as the model output $\boldsymbol{y(\theta)}$ 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 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 5 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 $\boldsymbol{z_{(xy)}}$ alone:

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

3.3 Emulation

For probabilistic projections 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 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 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 1×1 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:

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

The combined $\boldsymbol{\Omega}$ is in the following called emulator and $\boldsymbol{\omega}(\boldsymbol{\theta})$ as well as the entries of the diagonal matrix $\boldsymbol{\Sigma}_\omega(\boldsymbol{\theta})$ follow from equation 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 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 $\boldsymbol{\Theta}$ 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:

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

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; $\boldsymbol{\varepsilon} = N(0, \quad \boldsymbol{\Sigma}_\varepsilon)$. 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 $\boldsymbol{\Theta}$ and observational error $\boldsymbol{e}$ to be mutually statistically independent and $\boldsymbol{e}$ to be spatially identically distributed with variance σ_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 $(\mathbf{B'}^T\mathbf{B'})^{-\mathbf{1}}$ 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.

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

Note that the existence of $\boldsymbol{\gamma}$ is an abstract concept, implying that it is only because of an error ε 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 inclusion of model discrepancy 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, of the discrepancy are 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 $\mathbf{\Sigma}_\varepsilon$ is diagonal with $diag(\mathbf{\Sigma}_\varepsilon) = (\sigma^2_{\varepsilon 1}, ..., \sigma^2_{\varepsilon k})^T$. 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 σ^2_{i95} so that 95% of the observational distribution $N(\hat{z}_i, \quad \sigma^2_{ei})$ 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 $\sigma^2_{\varepsilon i} = MAX(\sigma^2_{i95}, \sigma^2_{ei})$.

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):

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

with $\mathbf{\Sigma}_T = \sigma^2_e (\mathbf{B'}^T \mathbf{B'})^{-\mathbf{1}} + \mathbf{\Sigma}_\varepsilon + \mathbf{\Sigma}_\omega$. 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:

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

The calibration distribution in Equation 6 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.

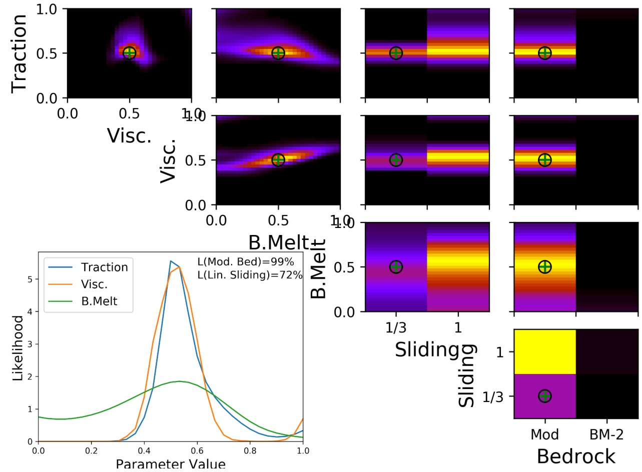

Figure 3. Likelihood of parameter combinations of synthetic test case (evaluations of Equation 12). Upper right panels show likelihood values marginalized to pairs of parameters, normalized to the respective maximum for clarity. Lower left panel shows likelihood values marginalized to individual parameters for the three scalar parameters (line plots), and sliding 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 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.

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 the 14 periods of satellite observations.

For Figure 3 the model run with central parameter values ($= 0.5$) for basal traction, viscosity and ocean melt scaling factors, nonlinear sliding 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 other synthetic model tests are shown in the supplement.

Figure 3 illustrates which parts of the model input space are most successful in reproducing the synthetic observations of surface elevation 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. 3 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.

As can be seen from Figure 3, marginal likelihoods of our calibration approach can favour linear sliding even if the synthetic observations use nonlinear sliding. 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 3 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 law and ocean melt overrules the weak constraints on these parameters in the marginalized likelihoods. The higher total likelihood of linear sliding can be traced back to a higher density of central ensemble members for linear sliding. Nonlinear sliding produces more extreme ice sheet simulations as fast simulations will have reduced (compared to linear sliding) basal drag and become even faster (and vice versa for slow simulations). The frequency distribution of total sea level contribution and basis representation are therefore wider for nonlinear sliding (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 compared to linear sliding (28% to 72%).

But why is the signal of sliding law and ocean melt 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? This is likely related to the delayed impact of those parameters compared to the others. 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 that the initial basal drag τ_b is approximately the same for both sliding laws with:

$$\tau_b = C_m(x,y) \cdot |v(x,y,t)|^{m-1} \cdot v(x,y,t) \tag{13}$$

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 is only after the ice velocities change that the sliding law has any impact on the simulations. A change in bedrock, basal traction or viscosity have, however, a much more immediate effect on the ice dynamics and are therefore expected to dominate the calibration on short time scales.

From this test we conclude that basal sliding law and ocean melt scaling cannot be inferred from this calibration approach. 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 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 law. It further is understood, that parts of the ASE bed consist of sediment-free, bare rocks for which a linear Weertman sliding law is not appropriate (Joughin et al., 2009). We therefore select nonlinear sliding 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 3 into context, we try two other methodical choices. First we calibrate in the spatial domain after re-projecting from the emulator results.

$$\boldsymbol{y'}(\boldsymbol{\theta}) = \mathbf{B}'\boldsymbol{\omega}(\boldsymbol{\theta}) \tag{14}$$

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 σ_e^2 so that the re-projected likelihood $L_{(xy)}$ simplifies to:

$$L_{(xy)}(\boldsymbol{z_{(xy)}}|\boldsymbol{\theta}) \propto \prod_{i=1}^{m} exp\left[-\frac{1}{2}\frac{(y'(\boldsymbol{\theta})_i - z_{(xy)i})^2}{3\sigma_e^2}\right] \tag{15}$$

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).

$$L_{SLC}(\boldsymbol{z_{(xy)}}|\boldsymbol{\theta}) \propto exp\left[-\frac{1}{2}\frac{(SLC(\boldsymbol{y'}(\boldsymbol{\theta})) - SLC(\boldsymbol{z_{(xy)}}))^2}{3\sigma_{SLC}^2}\right] \tag{16}$$

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. $\sigma_{SLC}^2 = VAR(SLC(\boldsymbol{z_{(xy)}})) = 0.035^2[\frac{mmSLE^2}{year^2}]$.

The calibration in (x,y) representation (Figure 4a) behaves similarly to the basis representation (Figure 3) in that sliding 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 but in addition, a wide range of traction-viscosity combinations perform equally well and there is no constraint on bedrock (Figure 4b). Furthermore, the model run used as synthetic observations is not identified as the most likely setup in Figure 4b. This demonstrates the value of the extra information - and stronger parameter constraints - provided by the use of two-dimensional observations.

4 Results

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

The calibration finds that the modified bedrock from Nias et al. (2016) produces much more realistic surface elevation changes than the original Bedmap2 topography (Fig. 5a). The weighted average of basal traction and velocity parameters are 0.47 and 0.45, respectively, which is slightly smaller the default values (0.5). This amounts to a 3.5% and 7.2% reduction in amplitude compared to the optimized fields from by (Nias et al., 2016).

We use the calibration in basis representation (likelihood shown in Fig. 5a) as well as the reprojected (x,y) and SLC based calibrations to update the projections of sea level contribution and grounding line retreat after 50 years in Figure 5b. As can

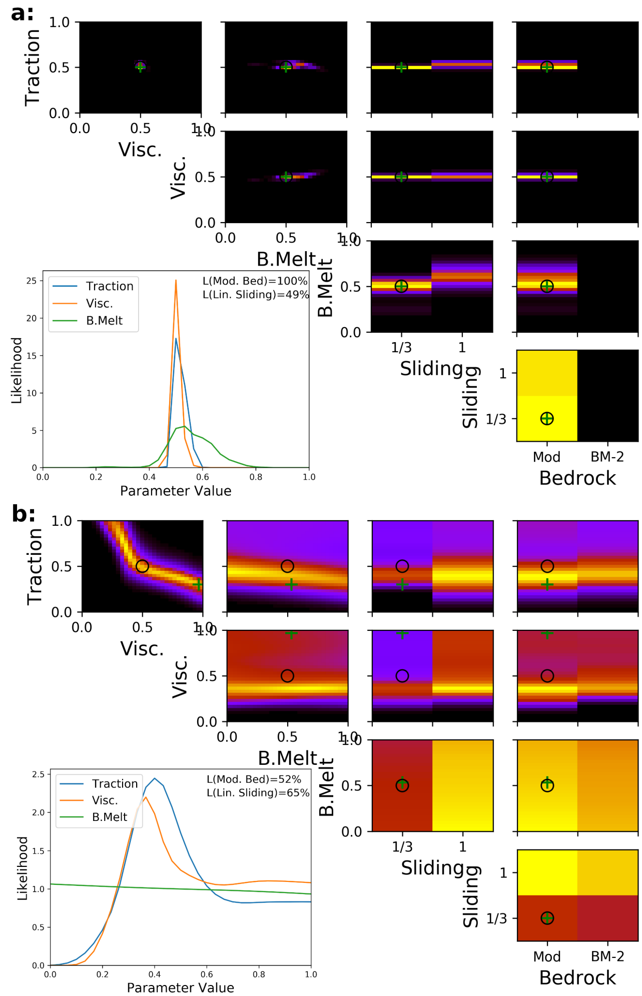

Figure 4. Likelihood of parameter combinations of synthetic test case for reprojected emulator estimates (top, a; Equation 15) and sea level rise contribution calibration (bottom, b; Equation 16). Upper right panels show likelihood values marginalized to pairs of parameters, normalized to the respective 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, consistent with Probability Density Functions. The central values for traction, viscosity and ocean melt as well as nonlinear sliding 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.

be seen from the Grey and Brown shaded histograms in Figure 5b (emulated and original BISICLES ensemble) 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 5b), as was indicated by the findings of Section 3.6. Calibration on the total sea level contribution

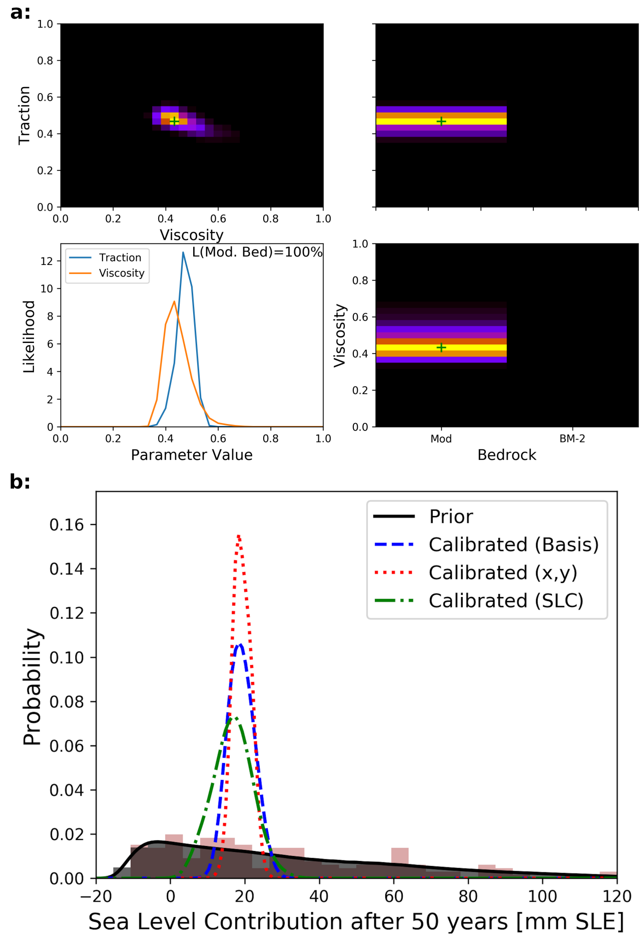

Figure 5. a: Likelihood of parameter combinations in basis representation from satellite observations (evaluations of Equation 12). Upper right panels show likelihood values marginalized to pairs of parameters, normalized to the respective maximum for clarity. Lower left panel shows likelihood values 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).

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 5b and Table 1)

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

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 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 ice sheet 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 laws (Nias et al., 2018). Here we find the modified bedrock topography to produce a spatial response far more consistent with observed surface elevation changes than for the original Bedmap2 bedrock (Fig. 5a). 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, 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 4b). Projections 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 4b). 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 4a 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. However, for one of the sub-periods with reprojected calibration the probability interval does not overlap with the results of the whole 7 year calibration period (Table 1 in the Supplement). Since the sub-period is part of the 7 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 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 ± 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 for basis, (x,y) and SLC, 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,, 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). Warmer ocean and air temperatures would enhance melt and accelerate the dynamic response. Neither do the used simulations 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.

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 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 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 Figure2). 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 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 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 variations to the input 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 model setups 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 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. 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 4 and 5b).

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. The temporal component could in particular help to constrain the basal sliding law exponent and ocean melt scaling.

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 from 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 to increase the amount of information used 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 law and ocean melt rate is more challenging, probably due to their slow impact on ice sheet behaviour compared to the short calibration timescale. 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 wide sea level contribution probability distributions. The extra information from the use of two-dimensional calibrations adds stronger parameter constrains, showing that this method can be used 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 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 50 year model period the Amundsen Sea Embayment is expected to contribute between 13.9 and 24.8 mm SLE (90% probability interval) with a most likely global sea level contribution of 18.4 mm SLE.

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.

References

Arthern, R. J. and Williams, C. R.: The sensitivity of West Antarctica to the submarine melting feedback, Geophysical Research Letters, 44, 2352–2359, 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, 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.

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, 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, 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, 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, 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, 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, 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, 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.

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, 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, 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.

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.

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, 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, 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.

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., 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.

MacAyeal, D. R., Bindschadler, R. A., and Scambos, T. A.: Basal friction of ice stream E, West Antarctica, Journal of Glaciology, 41, 247–262, 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, 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, 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, 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, 2016.

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

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.

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, 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, 2016.

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. M.: Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise, Geophysical Research Letters, 38, https://doi.org/10.1029/2011GL046583, 2011.

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, 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.

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, 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, 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.

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.

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, 2018.