

IMPROVING ICE SHEET MODEL CALIBRATION USING PALEOCLIMATE AND MODERN DATA

BY WON CHANG^{1,2,*}, MURALI HARAN^{1,3,†}, PATRICK
APPLEGATE^{1,3,†} AND DAVID POLLARD^{1,3,4,†}

University of Cincinnati^{*} and *Pennsylvania State University*[†]

Human-induced climate change may cause significant ice volume loss from the West Antarctic Ice Sheet (WAIS). Projections of ice volume change from ice sheet models and corresponding future sea-level rise have large uncertainties due to poorly constrained input parameters. In most future applications to date, model calibration has utilized only modern or recent (decadal) observations, leaving input parameters that control the long-term behavior of WAIS largely unconstrained. Many paleo-observations are in the form of localized time series, while modern observations are non-Gaussian spatial data; combining information across these types poses nontrivial statistical challenges. Here we introduce a computationally efficient calibration approach that utilizes both modern and paleo-observations to generate better constrained ice volume projections. Using fast emulators built upon principal component analysis and a reduced dimension calibration model, we can efficiently handle high-dimensional and non-Gaussian data. We apply our calibration approach to the PSU3D-ICE model which can realistically simulate long-term behavior of WAIS. Our results show that using paleo-observations in calibration significantly reduces parametric uncertainty, resulting in sharper projections about the future state of WAIS. One benefit of using paleo-observations is found to be that unrealistic simulations with overshoots in past ice retreat and projected future regrowth are eliminated.

1. Introduction. Human-induced climate change may cause significant ice volume loss in the polar regions. The West Antarctic Ice Sheet (WAIS) is particularly vulnerable to the changing climate because much of the ice is grounded well below sea level. Previous studies suggest that ice volume loss in this area may result in up to 4 meters of sea level rise [Fretwell et al. (2013)], which in turn will require significant resource allocations for adaptation and may therefore have a profound impact on human society [e.g., Lempert, Srivier and Keller (2012)].

Received May 2016; revised August 2016.

¹Supported in part by the NSF through Network for Sustainable Climate Risk Management under NSF cooperative agreement GEO1240507.

²Supported in part by the NSF Statistical Methods in the Atmospheric Sciences Network (Award Nos. 1106862, 1106974 and 1107046).

³Supported in part by the NSF through NSF-DMS-1418090.

⁴Supported in part by the NSF through NSF/OCE/FESD 1202632 and NSF/OPP/ANT 1341394.

Key words and phrases. Paleoclimate, West Antarctic Ice Sheet, computer model calibration, Gaussian process, dimension reduction.

Many recent modeling efforts have focused on simulating and projecting the past and future WAIS volume dynamics [e.g., [Bindshadler et al. \(2013\)](#)]. The relevant time scale for many aspects of WAIS behavior is often in hundreds to thousands of years; this in turn necessitates simulating and projecting the evolution of WAIS for a very long time span [[Cornford et al. \(2015\)](#), [Feldmann and Levermann \(2015\)](#), [Golledge et al. \(2015\)](#), [Gomez, Pollard and Holland \(2015\)](#), [Ritz et al. \(2015\)](#), [Winkelmann et al. \(2015\)](#)].

In this work we use the PSU3D-ICE model [[Pollard and DeConto \(2009, 2012a, 2012b\)](#)] to simulate the long-term evolution of the West Antarctic Ice Sheet. Using a hybrid dynamical core that combines the shallow-ice and the shallow-shelf approximation, the model can realistically simulate the long-term behavior of the ice sheet with a reasonable amount of computational effort. In contrast to higher-resolution models [e.g., [Favier et al. \(2014\)](#), [Gladstone et al. \(2012\)](#), [Joughin, Smith and Medley \(2014\)](#)] that are designed for relatively short simulations with more detailed physical processes, this modeling strategy enables us to simulate the long-term evolution of the West Antarctic Ice Sheet and utilize information from paleo-observations for model calibration. The approach here is a trade-off between (i) reduced fidelity in capturing details such as sills near modern grounding lines that may be important for 10^3 -km scale retreat in the next hundred years, and (ii) more robust calibration versus retreat over hundreds of kilometres and more pronounced bedrock variations, which is arguably more relevant to larger-scale retreat into the West Antarctic interior within the next several centuries.

Parametric uncertainty is an important source of uncertainty in projecting future WAIS volume change. Ice sheet models have input parameters that strongly affect the model behaviors; they are also poorly constrained [[Applegate et al. \(2012\)](#), [Stone et al. \(2010\)](#)]. Various calibration methods have been proposed to reduce parametric uncertainty for Greenland [[Chang et al. \(2014b\)](#), [McNeall et al. \(2013\)](#)] and the Antarctica ice sheets model [[Chang et al. \(2016\)](#), [Gladstone et al. \(2012\)](#)]. Although these recent studies have provided statistically sound ways of generating constrained future projections, they are mostly limited to generating short-term projections (i.e., a few hundred years from present) or utilizing modern or recent observations in the calibration. Inferring input parameters that are related to the long-term behavior of WAIS is crucial for generating well-constrained projections in the relevant time scale (hundreds to thousands of years). Modern or recent observations often lack information on these parameters, and therefore calibrating solely based on these information sources may result in poorly constrained projections. Recent studies using heuristic approaches suggest that utilizing information from paleo data can reduce uncertainties in these long-term behavior related parameters [[Briggs, Pollard and Tarasov \(2013, 2014\)](#), [Briggs and Tarasov \(2013\)](#), [Golledge et al. \(2014\)](#), [Maris et al. \(2015\)](#), [Whitehouse, Bentley and Le Brocq \(2012\)](#), [Whitehouse et al. \(2012\)](#)].

Here we propose an approach to simultaneously utilize modern- and paleo-observations for ice sheet model calibration, generating well-constrained future

WAIS ice volume change projections. This work represents the first statistically rigorous approach for calibrating an ice sheet model based on both modern and paleo data, and the resulting inference about parameters and projected sea level rise are therefore less uncertain than those obtained solely based on modern data [cf. Chang et al. (2016)]. Our methodological contribution is to provide a computationally expedient approach to fuse information from modern and paleo data. Our dimension reduction methods build upon two different calibration approaches given by Chang et al. (2014a, 2016), while also accounting for potential relationships between the two very different types of data—the modern data are spatial and binary, while the paleo data are in the form of a time series. A central contribution of this work is scientific. Based on our methods, we are able to show explicitly how paleo data provides key new information about parameters of the ice sheet model, and we thereby show that utilizing paleo data in addition to modern ice sheet data virtually eliminates the possibility of zero (or negative) sea level rise.

The rest of the paper is organized as follows. Section 2 introduces the model runs and the observational data sets used in our calibration experiment. Section 3 explains our computationally efficient reduced-dimension calibration approach that enables us to emulate and calibrate the PSU3D-ICE model using both the grounding line positions and the modern observations while avoiding the computational and inferential challenges. Section 4 describes our calibration experiment results based on a simulated example and real observations. Finally, in Section 5, we summarize our findings and discuss caveats and possible improvements.

2. Model runs and observational data. In this section, we describe the ice sheet model that we use to simulate past and future West Antarctic Ice Sheet behavior, as well as the modern and paleo-data sets that we use to calibrate the ice sheet model.

2.1. Model runs and input parameter description. We calibrate the following four model parameters that are considered to be important in determining the long-term evolution of the West Antarctic Ice Sheet, yet whose values are particularly uncertain: the sub-ice-shelf oceanic melt factor (OCFAC), the calving factor (CALV), the basal sliding coefficient (CRH), and the asthenospheric relaxation e -folding time (TAU). OCFAC (nondimensional) represents the strength of oceanic melting at the base of floating ice shelves in response to changing ocean temperatures, and CALV (nondimensional) determines the rate of iceberg calving at the outer edges of floating ice shelves. CRH ($\text{m year}^{-1} \text{Pa}^{-2}$) represents how slippery the bedrock is in areas around Antarctica that are currently under ocean, that is, how fast grounded ice slides over these areas as it expands beyond the present extent. A higher value of CRH corresponds to faster basal sliding and greater ice flux toward the ocean. TAU (with units in thousands of years) represents the time scale for vertical bedrock displacements in response to changing ice load. In this paper,

we remap the parameter values to the $[0, 1]$ intervals for convenience. We refer to Chang et al. (2016) for more detailed description of these parameters.

We run the PSU3D-ICE model with 625 different parameter settings specified by a factorial design, with five different values for each parameter. Starting from 40,000 years before present, each model run is spun up until present and then projected 5000 years into the future. For atmospheric forcing, we use the modern climatological Antarctic data set from the Sea Rise project [Bindshadler et al. (2013)] uniformly perturbed in proportion to a deep-sea-core $\delta^{18}\text{O}$ record [Pollard and DeConto (2009, 2012b)]. For oceanic forcing, we use the ocean temperature pattern from AOGCM simulation runs generated by Liu et al. (2009). From each model run we extract the following two data sets and compare them to the corresponding observational record: (i) the time series of grounding line positions, the location of the transition from grounded ice to ice shelf along the central flow-line in the Amundsen Sea Embayment (ASE) region (see Section 2.2 below for more details), and (ii) the modern binary spatial pattern of presence and absence of grounded ice in the ASE. The grounding line position time series (i) has 1501 time points from 15,000 years ago to the present, and the modern binary spatial pattern (ii) is a binary map with 86×37 pixels with a 20 km horizontal resolution. Because the time series of the grounding line position for each model run does not usually show much change until 15,000 years before present, we only use the time series after 15,000 years ago for our calibration. Note that each model output is in the form of high-dimensional multivariate data, which causes computational and inferential challenges described in Section 3.1.4. The corresponding observational data sets, described below, have the same dimensionalities as the model output sets.

Note that, for some parameter settings, the past grounding line position shows unrealistically rapid retreat early in the time series and does not change for the rest of the time period. We have found that using these runs for building our emulator negatively affects the emulation performance for more realistic model outputs. Therefore, we have excluded the model runs that reach a grounding line position 500 meters inland from the modern grounding line before 10,000 years ago from our analysis and use the remaining 461 model runs for emulating the past grounding line position output.

2.2. Paleo-records of grounding line positions. We take advantage of a very recent, comprehensive synthesis of Antarctic grounding line data since the last glacial maximum [RAISED Consortium (2014)]. For the ASE sector, Larter et al. (2014) provide spatial maps of estimated grounding lines at 5000 year intervals from 25,000 years ago to the present. These maps are based primarily on many ship-based observations taken in the oceanic part of ASE using sonar (showing patterns of ocean-floor features formed by flow of past grounded ice) and shallow sediment cores (providing dating and information on ice proximity, i.e., open

ocean, ice proximal or grounded). There is considerable uncertainty in the reconstructions, but general consensus for the overall retreat in this sector, especially along the central flowline of the major paleo-ice stream emerging from the confluence of Pine Island and Thwaites Glaciers and crossing the ASE [cf. earlier synthesis by Kirshner et al. (2012)].

2.3. Modern observations. We use a map of modern grounding lines deduced from the Bedmap2 data set [Fretwell et al. (2013)]. Bedmap2 is the most recent all-Antarctic data set that provides gridded maps of ice surface elevation, bedrock elevation and ice thickness. These fields were derived from a variety of sources, including satellite altimetry, airborne and ground radar surveys, and seismic sounding. The nominal Bedmap2 grid spacing is 1 km, but the actual coverage in some areas is sparser especially for ice thickness. We deduce grounding line locations by a simple floatation criterion at each model grid cell, after interpolating the data to our coarser model grid. In this work, we use a part of the data that corresponds to the ASE sector, which is expected to be the largest contributor to future sea level rise caused by WAIS volume loss [Pritchard et al. (2012)]. Since the observed modern binary pattern is derived from the highly accurate ice thickness measurements in the Bedmap2 data set and the model binary patterns are highly variable depending on the input parameter settings (see Section S1 and Figure S1 in the Supplementary Material [Chang et al. (2016)]), the model outputs approximated by our emulator are accurate enough to provide a basis for calibration.

3. Computer model emulation and calibration using dimension reduction.

As explained in the preceding discussion, parameter inference is central to WAIS volume change projections; taking full advantage of all observational data, both paleo and modern, may result in reduced parametric uncertainty which in turn can result in well-constrained projections of volume change. In this section we describe our statistical approach for inferring input parameters for WAIS models while accounting for relevant sources of uncertainty. In Section 3.1 we introduce our two-stage framework [Bayarri et al. (2007), Bhat, Haran and Goes (2010), Bhat et al. (2012), Chang et al. (2014a)] that consists of the emulation and the calibration steps: In the emulation step we build a Gaussian process emulator as a fast approximation to WAIS model outputs [Sacks et al. (1989)]. In the calibration step, we infer the input parameters for WAIS models by combining information from emulator output and observational data while accounting for the systematic model-observation discrepancy. The framework faces computational and inferential challenges when model output and observational data are high-dimensional data such as large spatial patterns or long time series, and the challenges are further exacerbated when the marginal distribution of model output and observational data cannot be modeled by a Gaussian distribution. Section 3.2 describes a computationally expedient reduced-dimension approach that mitigates these challenges

for high-dimensional Gaussian and non-Gaussian data and enables us to utilize information from both past grounding line positions and modern binary patterns for calibration.

We use the following notation henceforth: Let $\theta_1, \dots, \theta_q \in R^4$ be the parameter settings at which we use both the past grounding line positions and the modern binary patterns for emulation, and let $\theta_{q+1}, \dots, \theta_p \in R^4$ be the settings at which we use only the modern binary patterns (see Section 2.1 above for the reason why q is less than p in our experiment). We denote the past grounding line position time series from our WAIS model at a parameter setting θ and a time point t by $Y_1(\theta, t)$. We let \mathbf{Y}_1 be a $q \times n$ matrix where its i th row is $[Y_1(\theta_i, t_1), \dots, Y_1(\theta_i, t_n)]$ and t_1, \dots, t_n are time points at which the grounding line positions are recorded. We let $\mathbf{Z}_1 = [Z_1(t_1), \dots, Z_1(t_n)]$ be a vector of the observed time series of past grounding line positions reconstructed from paleo records. Similarly, we denote the modern ice-no ice binary output at the parameter setting θ and a spatial location \mathbf{s} by $Y_2(\theta, \mathbf{s})$. We let \mathbf{Y}_2 be a $p \times m$ matrix where its i th row is $[Y_2(\theta_i, \mathbf{s}_1), \dots, Y_2(\theta_i, \mathbf{s}_m)]$ with model grid points $\mathbf{s}_1, \dots, \mathbf{s}_m$. The corresponding observational data are denoted by an m -dimensional vector $\mathbf{Z}_2 = [Z_2(\mathbf{s}_1), \dots, Z_2(\mathbf{s}_m)]$. For our WAIS model emulation and calibration problem in Section 4, $n = 1501$, $m = 3182$, $p = 625k$ and $q = 461$.

3.1. *Basic WAIS model emulation and calibration framework.* In this subsection we explain the basic general framework for computer model emulation and calibration, and describe the computational challenges posed by our use of high-dimensional model output and observational data.

3.1.1. *Emulation and calibration using past grounding line positions.* We start with the model output \mathbf{Y}_1 and the observational data \mathbf{Z}_1 for the past grounding line positions. Since computer model runs are available only at a limited number of parameter settings q , one needs to construct a statistical model for approximating the model output at a new parameter setting θ by interpolating the existing model output obtained at the design points $\theta_1, \dots, \theta_q$ [Sacks et al. (1989)]. Constructing this statistical model requires us to build a Gaussian process that gives the following probability model for the existing q model runs with n -dimensional output:

$$(1) \quad \text{vec}(\mathbf{Y}_1) \sim N(\mathbf{X}_1 \boldsymbol{\beta}_1, \boldsymbol{\Sigma}(\boldsymbol{\xi}_1)),$$

where $\text{vec}(\cdot)$ is the vectorization operator that stacks the columns of a matrix into one column vector, and \mathbf{X}_1 is a $nq \times b$ covariate matrix that contains all the time coordinates and the input parameters settings used in the $nq \times nq$ covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\xi}_1)$ with a covariance parameter vector $\boldsymbol{\xi}_1$. The b -dimensional vector $\boldsymbol{\beta}_1$ contains all the coefficients for the columns of \mathbf{X}_1 . When the number of time points n and the number of parameter settings q are small, one can estimate the parameter $\boldsymbol{\xi}_1$ by maximizing the likelihood function corresponding to the probability model in (1) and finding the conditional distribution of $Y_1(\theta_i, t_1), \dots, Y_1(\theta_i, t_n)$ given \mathbf{Y}_1

for any new value of θ using the fitted Gaussian process with the maximum likelihood estimates $\hat{\beta}_1$ and $\hat{\xi}_1$. We call the fitted Gaussian process an emulator and denote the output at θ interpolated by the emulator as $\eta(\theta, \mathbf{Y}_1)$. Using the emulator, one can set up a model for inferring the input parameter θ as follows [Bayarri et al. (2007), Kennedy and O’Hagan (2001)]:

$$(2) \quad \mathbf{Z}_1 = \eta(\theta, \mathbf{Y}_1) + \delta,$$

where δ is an n -dimensional random vector that represents model-observation discrepancy. The discrepancy term δ is often modeled by a Gaussian process that is independent of the emulated output $\eta(\theta, \mathbf{Y}_1)$. Using a posterior density defined by the likelihood function that corresponds to the probability model in (2) and a standard prior specification, one can estimate the input parameter θ along with other parameters in the model via Markov Chain Monte Carlo (MCMC).

3.1.2. *Emulation and calibration using modern binary observations.* Emulation and calibration for the modern ice-no ice binary patterns require additional consideration in model specification due to the binary nature of the data sets. Inspired by the generalized linear model framework, Chang et al. (2016) specify emulation and calibration models in terms of logits of model output and observational data. Let $\Gamma = \{\gamma_{ij}\}$ be a $p \times m$ -dimensional matrix whose element is the logit of the (i, j) th element in \mathbf{Y}_2 , that is,

$$\begin{aligned} P(Y_2(\theta_i, \mathbf{s}_j) = y_{ij}) &= \left(\frac{\exp(\gamma_{ij})}{1 + \exp(\gamma_{ij})} \right)^{y_{ij}} \left(\frac{1}{1 + \exp(\gamma_{ij})} \right)^{1-y_{ij}} \\ &= (1 + \exp(-(2y_{ij} - 1)\gamma_{ij}))^{-1}, \end{aligned}$$

where the value of y_{ij} is either 0 or 1. Assuming conditional independence between the elements in \mathbf{Y}_2 given Γ , one can use a Gaussian process that yields the probability model below to specify the dependence between the model output at different parameter settings and spatial locations,

$$(3) \quad \text{vec}(\Gamma) \sim N(\mathbf{X}_2\beta_2, \Sigma(\xi_2)),$$

where the $mp \times c$ dimensional covariate matrix \mathbf{X}_2 , c -dimensional coefficient vector β_2 and the $mp \times mp$ covariance matrix $\Sigma(\xi_2)$ with a covariance parameter vector ξ_2 are defined in the same way as in (1). One can find the maximum likelihood estimates $\hat{\beta}_2$ and $\hat{\xi}_2$ by maximizing the likelihood function corresponding to the probability model in (3). The resulting Gaussian process emulator gives a vector of interpolated logits $\psi(\theta, \mathbf{Y}_2)$ for a new input parameter value θ .

The calibration model is also defined in terms of the logits of the observational data \mathbf{Z}_2 , denoted by an m -dimensional vector $\lambda = [\lambda_1, \dots, \lambda_m]$,

$$\lambda = \psi(\theta, \mathbf{Y}_2) + \phi,$$

where ϕ is an m -dimensional random vector representing the model-observation discrepancy defined in terms of the logits of \mathbf{Z}_2 . Again, assuming conditional independence between the elements in \mathbf{Z}_2 given λ , the relationship between λ and $Z_2(\mathbf{s}_i)$ is given by

$$(4) \quad P(Z_2(\mathbf{s}_j) = z_j) = \left(\frac{\exp(\lambda_j)}{1 + \exp(\lambda_j)} \right)^{z_j} \left(\frac{1}{1 + \exp(\lambda_j)} \right)^{1-z_j} \\ = (1 + \exp(-(2z_j - 1)\lambda_j))^{-1},$$

where z_j takes a value of either 0 or 1. If the number of parameter settings p and spatial locations m are small, one can set up a posterior density based on the likelihood function corresponding to the probability models above and some standard prior specifications, and might be able to infer θ and other parameters using MCMC.

3.1.3. Combining information from two data sets in calibration. We set up a calibration model to infer the input parameters in θ based on the models described in Sections 3.1.1 and 3.1.2. The main consideration here is how to model the dependence between \mathbf{Z}_1 and \mathbf{Z}_2 , which can be translated to the dependence between the emulated outputs $\eta(\theta, \mathbf{Y}_1)$ and $\psi(\theta, \mathbf{Y}_2)$ and the dependence between the discrepancy terms δ and ϕ . We model the dependence between $\eta(\theta, \mathbf{Y}_1)$ and $\psi(\theta, \mathbf{Y}_2)$ only through the input parameter θ [i.e., we assume that $\eta(\theta, \mathbf{Y}_1)$ and $\psi(\theta, \mathbf{Y}_2)$ are conditionally independent given the input parameter θ] because the emulators are independently constructed for \mathbf{Y}_1 and \mathbf{Y}_2 . We do not introduce a conditional dependence between the emulators $\eta(\theta, \mathbf{Y}_1)$ and $\psi(\theta, \mathbf{Y}_2)$ given θ because the emulators are already highly accurate. The greater challenge is the dependence between the discrepancy terms δ and ϕ because both terms are high dimensional and it is not straightforward to find a parsimonious model that can efficiently handle the cross-correlation between them (see Section 3.1.4 below for further discussion and Section 3.2 for our solution).

3.1.4. Computational and inferential challenges. The emulation and calibration problems for both the past grounding line positions and the modern ice-no ice binary patterns face computational and inferential challenges when the length of time series n and the number of spatial locations m are large. For both of these problems, the likelihood evaluation in the emulation step involves Cholesky decomposition of $nq \times nq$ and $mp \times mp$ covariance matrices, which scales as $\mathcal{O}(n^3q^3)$ and $\mathcal{O}(m^3p^3)$, respectively. For our WAIS model calibration problem, this requires $\frac{1}{3}n^3q^3 = 1.1 \times 10^{17}$ and $\frac{1}{3}m^3p^3 = 2.6 \times 10^{18}$ flops of computation for each likelihood evaluation, which translate to about 28,000 hours and 220,000 hours on a high-performance single core. Moreover, emulation and calibration using the modern ice-no ice patterns poses additional inferential difficulties. In particular, we need to compute $mp = 1,988,750$ logits for the model output in the

emulation step and $2m = 6364$ logits for the observational data. The challenge is therefore to ensure that the problem is well posed by constraining the logits, while at the same time retaining enough flexibility in the model. The problem is even more complicated due to the dependence between the discrepancy terms δ and ϕ because we need to estimate $n \times m = 1501 \times 3182$ correlation coefficients between the elements in those two terms.

3.2. Reduced-dimension approach. In this subsection, we discuss our reduced-dimension approaches to mitigate the computational and inferential challenges described above.

3.2.1. Dimension reduction using principal components. The first step is to reduce the dimensionality of model output via principal component analysis. For the model output matrix of the past grounding line positions \mathbf{Y}_1 , we find the J_1 -leading principal components by treating each of its columns (i.e., output for each time point) as different variables and rows (i.e., output for each parameter setting) as repeated observations. For computational convenience we rescale the principal component scores by dividing them by the square roots of their corresponding eigenvalues so that their sample variances become 1. We denote the j th rescaled principal component scores at the parameter setting θ as $Y_1^R(\theta, j)$, and denote the $q \times J_1$ matrix that contains all the principal component scores for the design points $\theta_1, \dots, \theta_q$ as $\mathbf{Y}_1^R = \{Y_1^R(\theta_i, j)\}$ with its rows for different parameter settings and columns for different principal components. Similarly, for the model output matrix of the modern ice-no ice binary patterns \mathbf{Y}_2 , we form a $p \times J_2$ matrix $\mathbf{Y}_2^R = \{Y_2^R(\theta_i, j)\}$ of J_2 leading logistic principal components in the same way, where $Y_2^R(\theta, j)$ is the j th logistic principal component at the parameter setting θ .

We use $J_1 = 20$ principal components for the past grounding position output and $J_2 = 10$ for the modern binary pattern output. Through a cross-validation experiment described below in Section 4, we have found that increasing the number of principal components does not improve the emulation performance. We have also confirmed that the principal component score surfaces vary smoothly in the parameter space, and hence Gaussian process emulation is a suitable approach to approximating them (Figures S2–S7 in the Supplementary Material [Chang et al. (2016)]).

We display the first three principal components for a modern binary spatial pattern in Figure S8 and past grounding line position time series in Figure S9. The first three principal components for the modern binary spatial pattern show that the most variable patterns between parameter settings are (i) the overall ice coverage in the inner part of the Amundsen Embayment, which determines whether there is a total collapse of ice sheet in this area, (ii) the grounding line pattern around the edge of Amundsen Sea Embayment, (iii) and the ice coverage around the Thurston island. The first three principal components for the past grounding

line position time series indicate that the most variable patterns between the input parameter settings are (i) the grounding line retreat occurring between 15,000 and 7000 years ago, (ii) the retreat occurring until around 9000 years ago, followed by strong re-advance until 3000 years ago (red dashed curve), and (iii) a quasi-sinusoidal advance and retreat (black dashed-dotted curve) that spans the entire time period.

3.2.2. Emulation using principal components. In the emulation step, our principal component-based approach allows us to circumvent expensive matrix computations in likelihood evaluation (Section 3.1.4 above) by constructing emulators for each principal component separately. For the j th principal component of \mathbf{Y}_1 , we fit a Gaussian process model to $Y_1^R(\boldsymbol{\theta}_1, j), \dots, Y_1^R(\boldsymbol{\theta}_q, j)$ ($j = 1, \dots, J_1$) with 0 mean and the following covariance function:

$$\text{Cov}(Y_1^R(\boldsymbol{\theta}_k, j), Y_1^R(\boldsymbol{\theta}_l, j)) = \kappa_{1,j} \exp\left(-\sum_{i=1}^4 \frac{|\theta_{ik} - \theta_{il}|}{\phi_{1,ij}}\right) + \zeta_{1,j} \mathbf{1}(\boldsymbol{\theta}_k = \boldsymbol{\theta}_l),$$

with $\kappa_{1,j}, \phi_{1,1j}, \dots, \phi_{1,4j}, \zeta_{1,j} > 0$ by finding the MLEs $\hat{\kappa}_{1,j}, \hat{\phi}_{1,1j}, \dots, \hat{\phi}_{1,4j}$, and $\hat{\zeta}_{1,j}$. The computational cost for likelihood evaluation is reduced from $\mathcal{O}(q^3 n_1^3)$ to $\mathcal{O}(J_1 q^3)$. The resulting J_1 Gaussian process models allow us to interpolate the values of the principal components at any new value of $\boldsymbol{\theta}$. We denote the collection of the predicted values given by these Gaussian process models for parameter setting $\boldsymbol{\theta}$ as $\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}_1^R)$. Similarly, we construct Gaussian process models for the logistic principal components $Y_2^R(\boldsymbol{\theta}_1, j), \dots, Y_2^R(\boldsymbol{\theta}_p, j)$ ($j = 1, \dots, J_2$) with mean 0 and the covariance function

$$\text{Cov}(Y_2^R(\boldsymbol{\theta}_k, j), Y_2^R(\boldsymbol{\theta}_l, j)) = \kappa_{2,j} \exp\left(-\sum_{i=1}^4 \frac{|\theta_{ik} - \theta_{il}|}{\phi_{2,ij}}\right) + \zeta_{2,j} \mathbf{1}(\boldsymbol{\theta}_k = \boldsymbol{\theta}_l),$$

with $\kappa_{2,j}, \phi_{2,1j}, \dots, \phi_{2,4j}, \zeta_{2,j} > 0$, by finding the MLEs $\hat{\kappa}_{2,j}, \hat{\phi}_{2,1j}, \dots, \hat{\phi}_{2,4j}$, and $\hat{\zeta}_{2,j}$. This reduces the computational cost for likelihood evaluation from $\mathcal{O}(m^3 p^3)$ to $\mathcal{O}(J_2 p^3)$. Moreover, our approach requires computing only $p \times J_2$ logistic principal components, and hence eliminates the need for computing mp logits. As above, we let $\boldsymbol{\psi}(\boldsymbol{\theta}, \mathbf{Y}_2^R)$ be the collection of the values of logistic principal components at any new value of $\boldsymbol{\theta}$ interpolated by the Gaussian process models.

3.2.3. Dimension-reduced calibration. In the calibration step, we use basis representations for the observational data sets using the emulators for the principal components constructed above to mitigate the computational and inferential challenges explained in Section 3.1.4. For the observed past grounding line positions \mathbf{Z}_1 , we set up the following linear model:

$$(5) \quad \mathbf{Z}_1 = \mathbf{K}_{1,y} \boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}_1^R) + \mathbf{K}_{1,d} \mathbf{v}_1 + \boldsymbol{\varepsilon}_1,$$

where $\mathbf{K}_{1,y}$ is the $n \times J_1$ matrix for the eigenvectors for the leading principal components rescaled by square roots of their corresponding eigenvalues, $\mathbf{K}_{1,d}\mathbf{v}_1$ is a low-rank representation of the discrepancy term δ , with an $n \times M$ basis matrix $\mathbf{K}_{1,d}$ and its M -dimensional random coefficient vector $\mathbf{v}_1 \sim N(\mathbf{0}, \alpha_1^2 \mathbf{I}_M)$ ($\alpha_1^2 > 0$), and $\boldsymbol{\varepsilon}_1$ is a vector of n i.i.d. random errors with mean 0 and variance $\sigma_\varepsilon^2 > 0$ (see Section 3.3 below for the details on specifying $\mathbf{K}_{1,d}$). Inferring parameters using dimension-reduced observational data computed based on the representation in (5) leads to a significant computational advantage by reducing the computational cost for likelihood evaluation from the order of $\mathcal{O}(n^3)$ to the order of $\mathcal{O}((J_1 + M)^3)$ (see Appendix A for details).

The idea of using principal components and kernel convolution in calibration is similar to the approach described in Higdon et al. (2008). However, our approach enables a faster computation by emulating each principal component separately and formulating the calibration model in terms of the dimension-reduced observational data \mathbf{Z}_1^R ; the approach in Higdon et al. (2008) retains the original data \mathbf{Z}_1 , and hence their computational gains are primarily due to more efficient matrix operations. Moreover, we use a two-stage approach [cf. Bayarri et al. (2007), Bhat et al. (2012), Chang et al. (2014a)], which separates the emulation and the calibration steps, to reduce the identifiability issues between the parameters in the emulator and the discrepancy term.

For the modern observed ice-no ice binary patterns \mathbf{Z}_2 , we set up the following linear model for the logits:

$$(6) \quad \boldsymbol{\lambda} = \mathbf{K}_{2,y}\boldsymbol{\psi}(\boldsymbol{\theta}, \mathbf{Y}_2^R) + \mathbf{K}_{2,d}\mathbf{v}_2,$$

where $\mathbf{K}_{2,y}$ is the $m \times J_2$ eigenvectors for the leading logistic principal components and $\mathbf{K}_{2,d}$ is an $m \times L$ basis matrix with L -dimensional random coefficients $\mathbf{v}_2 \sim N(\mathbf{0}, \alpha_2^2 \mathbf{I}_L)$ (see Section 3.3 below for the details on specifying $\mathbf{K}_{2,d}$). This basis representation also reduces the cost for matrix computation from $\mathcal{O}(m^3)$ to $\mathcal{O}(J_2 p^3)$. More importantly, using this basis representation reduces the number of logits that need to be estimated from $2m$ to $J_2 + L$, and hence makes the calibration problem well posed. Using the model in (5) and (6) and additional prior specification, we can set up the posterior density and estimate the input parameters $\boldsymbol{\theta}$ while accounting for the uncertainty in other parameters via MCMC using the standard Metropolis–Hastings algorithm. We describe posterior density specification in more detail in Appendix B.

We need to consider the dependence between \mathbf{Z}_1 and \mathbf{Z}_2 to use the information from both data sets simultaneously. As discussed above, we model the dependence between the emulators through the input parameter $\boldsymbol{\theta}$. We also capture the dependence between the discrepancy terms through the $M \times L$ cross-correlation matrix \mathbf{R}_v between \mathbf{v}_1 and \mathbf{v}_2 , where the (i, j) th element of \mathbf{R}_v is $\rho_{v,ij} = \text{Cov}(v_{1i}, v_{2j})$ and v_{1i} and v_{2j} are respectively the i th and j th elements of \mathbf{v}_1 and \mathbf{v}_2 (see Appendix A and B for further details). This greatly reduces the inferential issue by reducing the number of cross-correlation coefficients that need to be estimated from mn to ML .

3.3. *Model-observation discrepancy.* For a successful computer model calibration it is important to find good discrepancy basis matrices $\mathbf{K}_{1,d}$ and $\mathbf{K}_{2,d}$ that allow enough flexibility in representing the model-observation discrepancies while avoiding identifiability issues in estimating θ [cf. Brynjarsdóttir and O'Hagan (2014)]. To define the discrepancy basis matrix for the past grounding line positions $\mathbf{K}_{1,d}$, we use a kernel convolution approach using $M < n$ knot points a_1, \dots, a_M that are evenly distributed between t_1 and t_n . We use the following exponential kernel function to define the correlation between t_1, \dots, t_n and a_1, \dots, a_M :

$$(7) \quad \{\mathbf{K}_{1,d}\}_{ij} = \exp\left(-\frac{|t_i - a_j|}{\phi_{1,d}}\right),$$

with a fixed value $\phi_{1,d} > 0$. The basis representation based on this kernel function enables us to represent the general trend in model-observation discrepancy using a small number of random variables for the knot locations. Note that the discrepancy term constructed by kernel convolution can confound the effect from model parameters and thus cause identifiability issues; any trend produced by $\mathbf{K}_{1,d}\eta(\theta, \mathbf{Y}_1^R)$ can be easily approximated by $\mathbf{K}_{1,d}\mathbf{v}_1$, and therefore one cannot distinguish the effects from these two terms [Chang et al. (2014a)]. To avoid this issue, we replace $\mathbf{K}_{1,d}$ with its leading eigenvectors, which corresponds to regularization given by ridge regression [see Hastie, Tibshirani and Friedman (2009), page 66]. In the calibration experiment in Section 4, we chose the value of $\phi_{1,d}$ as 750 (years), the number of knots as $M = 1500$ and the number of eigenvectors as 300, and confirmed that using the discrepancy term based on these values leads to a reasonable calibration result by a simulated example (Section 4.1). We have also found that different settings for these values lead to similar calibration results, and hence inference for θ is robust to the choice of these values [cf. Chang et al. (2014a)].

The identifiability issue explained above is further complicated for the modern ice-no ice binary patterns because binary patterns provide even less information for separating the effects from the input parameters and the discrepancy term than continuous ones do. Through some preliminary experiments (not shown here) we found that the regularization introduced above does not solve the identifiability issue for binary patterns. Therefore, we use an alternative approach to construct the discrepancy basis $\mathbf{K}_{2,d}$, which is based on comparison between model runs and observational data [Chang et al. (2016)]. In this approach $\mathbf{K}_{2,d}$ has only one column (i.e., $L = 1$), and therefore the matrix is reduced to a column vector $\mathbf{k}_{2,d}$ and its coefficient vector \mathbf{v}_2 becomes a scalar v_2 . For the j th location \mathbf{s}_j , we calculate the following signed mismatch between the model and observed binary outcomes:

$$r_j = \frac{1}{P} \sum_{i=1}^P \text{sgn}(Y_2(\theta_i, \mathbf{s}_j) - Z_2(\mathbf{s}_j)) I(Y_2(\theta_i, \mathbf{s}_j) \neq Z_2(\mathbf{s}_j)),$$

where $\text{sgn}(\cdot)$ is the sign function. If $|r_j|$ is greater than or equal to a threshold value c , then we identify \mathbf{s}_j as a location with notable discrepancy and define the corresponding j th element of $\mathbf{k}_{2,d}$ as the logistic transformed r_j , $\log\left(\frac{1+r_j}{1-r_j}\right)$. If $|r_j| < c$, then we assume that the location \mathbf{s}_j shows no notable model-observation discrepancy and set the j th element of $\mathbf{k}_{2,d}$ as 0. Choosing a too large value of c results in inaccurate discrepancy representation by ignoring important patterns in the model-observation discrepancy, while a too small value of c causes identifiability issues between the input parameters $\boldsymbol{\theta}$ and the discrepancy term. Based on experiments with different model runs and observational data sets [cf. Chang et al. (2016)], we found that setting c to be 0.5 gives us a good balance between accurate discrepancy representation and parameter identifiability.

4. Implementation details and results. We calibrate the PSU3D-ICE model (Section 2) using our reduced-dimension approach (Section 3.2). We first verify the performance of our approach using simulated observational data sets (Section 4.1) and then move on to actual calibration using real observational data sets to estimate the input parameters and generate WAIS volume change projections (Section 4.2).

Before calibration, we verified the performance of our emulators through separate leave-out experiments for each emulator η and ψ . In each experiment we leave out a group of model runs around the center of the parameter space from the ensemble and try to recover them using an emulator constructed based on the remaining model runs. We have left out 82 model runs for the modern binary ice patterns and 60 for the past grounding line positions since we use a smaller number of model runs to emulate the grounding line position output ($q = 461$) than the modern binary pattern output ($p = 625$). Some examples of the comparison results are shown in Figures 1 (for the past grounding line positions) and 2 (for the modern binary patterns). The results show that our emulators can approximate the true model output reasonably well.

4.1. *Simulated example.* In this subsection we describe our calibration results based on simulated observations to study how accurately our method recovers the assumed true parameter setting and its corresponding true projected ice volume change. The assumed-true parameter setting that we choose to use here is $\text{OCFAC} = 0.5$, $\text{CALV} = 0.5$, $\text{CRH} = 0.5$ and $\text{TAU} = 0.4$ (rescaled to $[0, 1]$), which correspond to $\text{OCFAC} = 1$ (nondimensional), $\text{CALV} = 1$ (nondimensional), $\text{CRH} = 10^{-7}$ (m/year Pa²) and approximately $\text{TAU} = 2.6$ (k year) in the original scale. This is one of the design points that is closest to the center of the parameter space.

To represent the presence of model-observation discrepancy, we contaminate the model outputs at the true parameter setting with simulated structural errors. We generate simulated errors for the past grounding line positions from a Gaussian process model with zero mean and the covariance defined by the squared exponential function with a sill of 90 m and a range of 10,500 years. The generated

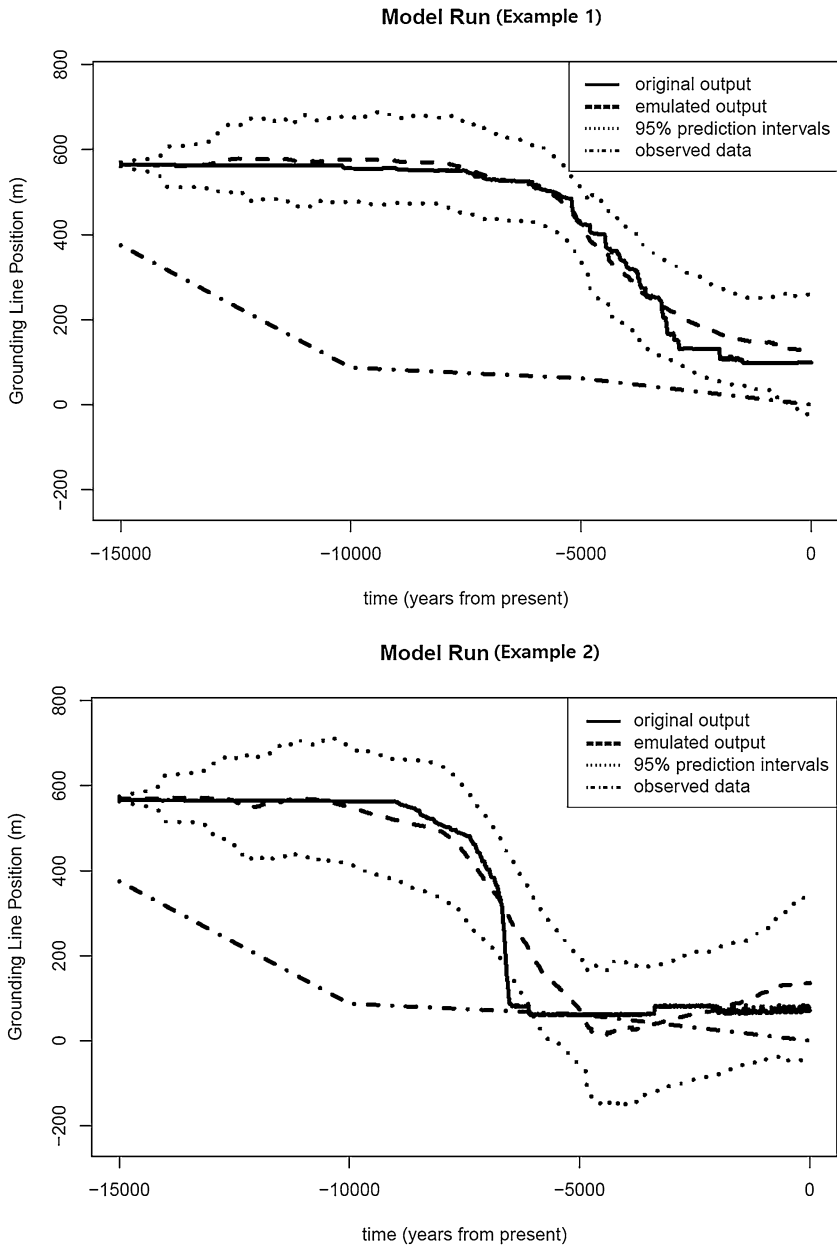


FIG. 1. Results for two example parameter settings from the leave-out experiment to verify the performance of the emulator for past grounding line positions. Results for other parameter settings are qualitatively similar to the results shown here. In general, the emulated grounding positions are similar to those from the actual model runs. For comparison we have also added the reconstructed grounding line position observations.

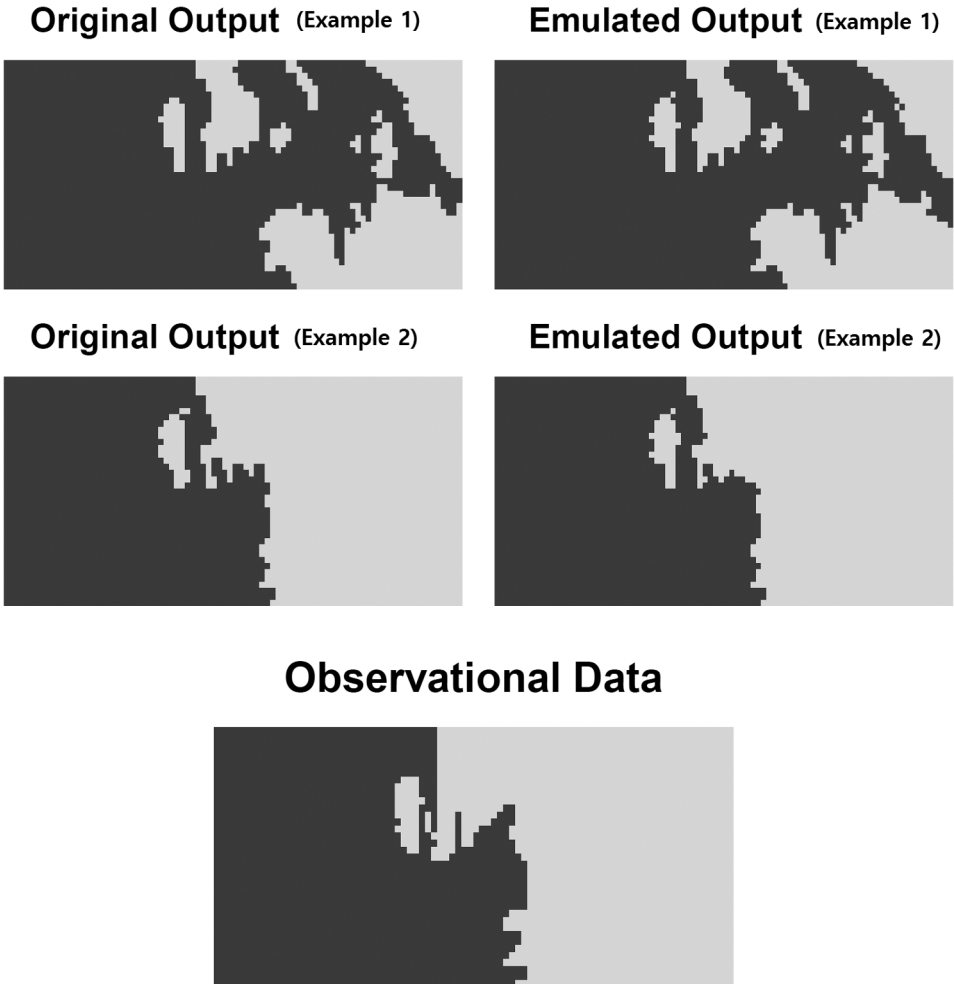


FIG. 2. Results for two example parameter settings from the leave-out experiment to verify the performance of the emulator for modern binary patterns (light gray for grounded ice and dark gray for no grounded ice). Results for other parameter settings are similar to the ones presented here. In general, the emulator can accurately approximate the binary patterns from the actual model runs. For comparison we have also included the observed modern binary pattern.

errors represent a situation where the model-observation discrepancy varies slowly and has a persistent trend over time. We generate the discrepancy for the modern binary patterns in a manner that makes them realistic. Our approach is therefore as follows. We use the original model output and observational data for the ice thickness pattern used to derive the modern binary patterns. We first choose the “best” 90% model runs (i.e., exclude the “worst” 10% model runs) that are closest to the modern observations in terms of the mean squared differences, and take pixel-wise averages for the selected model runs to derive a common thickness pat-

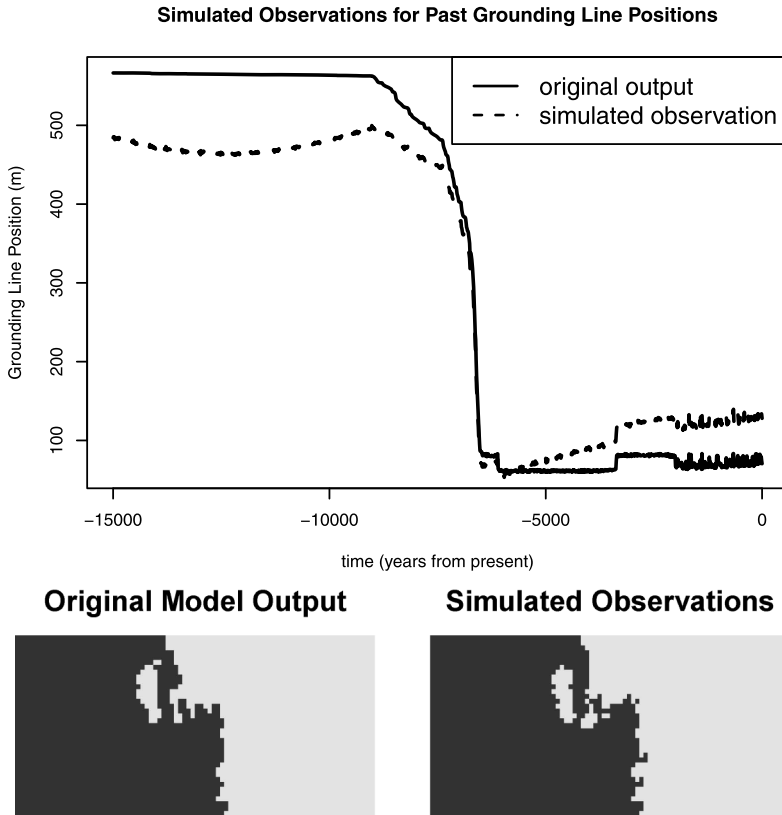


FIG. 3. The original model outputs at the assumed true parameter setting and the simulated observations used for the simulated example in Section 4.1. The upper plot is for the past grounding line positions and the lower plots are for the modern binary patterns (light gray for grounded ice and dark gray for no grounded ice).

tern. We then subtract the common thickness pattern from the observational data to derive the common discrepancy pattern in terms of the ice thickness. By subtracting this discrepancy pattern from the ice thickness pattern for the assumed truth and dichotomizing the resulting thickness pattern into a binary ice-no ice pattern, we obtain simulated observations for the modern binary pattern. We illustrate the resulting simulated observations in Figure 3.

Depending on the computational environment, the emulation step typically takes several minutes to half an hour. The computation in the calibration step requires about 240 hours to obtain an MCMC sample of size 70,000. The values of MCMC standard errors [Flegal, Haran and Jones (2008), Jones et al. (2006)] suggest that the sample size is large enough to estimate the posterior mean. We have also compared the marginal densities estimated from the first half the MCMC chain and the whole chain and confirmed that the chain is stable.

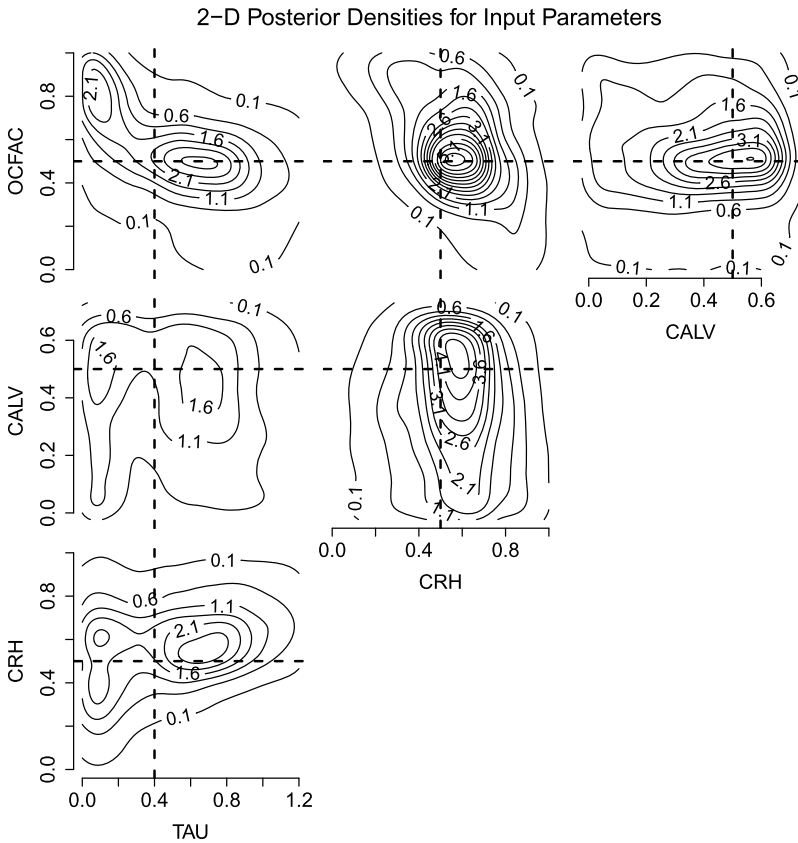


FIG. 4. Posterior density for input parameters from the perfect model experiment described in Section 4.1. The black dashed lines represent the assumed parameter settings. The result indicates that our method can recover the true input parameter setting by assigning high posterior density around the truth. The joint density in the upper left panel shows that OCFAC and TAU have a strong nonlinear interaction.

The probability density plots in Figure 4 show that the posterior has a high probability mass around the assumed truth, and hence indicate that our approach can recover the true parameter setting. However, the results also show that parameter settings other than the assumed truth have high probability densities, suggesting that settings other than the assumed truth can yield similar outputs to the simulated observations. Interestingly, the joint posterior density for TAU and OCFAC shows that these two parameters have a strong nonlinear interaction with each other.

Figure 5 shows the resulting projection for ice volume change for 500 years from present. To generate the projections, we first built an emulator to interpolate the ice volume change values at the existing 625 parameter settings and then converted our MCMC chain for θ into a Monte Carlo sample for the ice volume

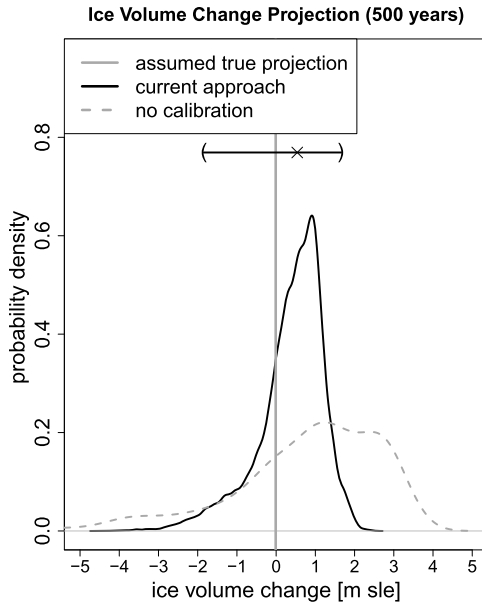
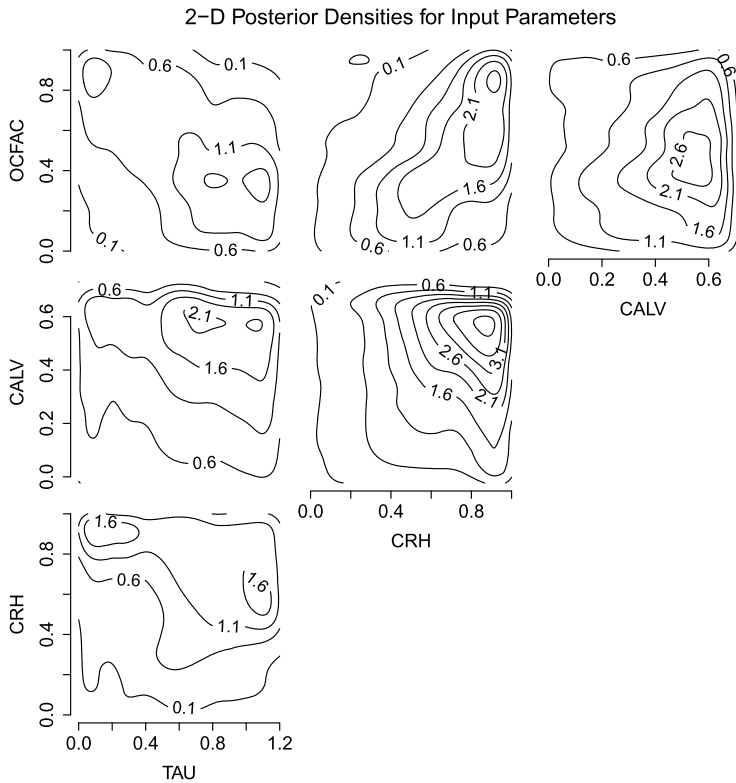


FIG. 5. Predictive distribution for the ice volume change projection based on the simulated observations described in Section 4.1. The gray solid line shows the assumed truth. The black solid and the gray dashed lines, respectively, show the predictive distributions with and without calibration using our approach. The result shows that our approach can recover the assumed truth and significantly reduce the uncertainty in projection.

change values using the emulator. The result also confirms that our method accurately recovers the assumed-true projection with reasonable uncertainty.

4.2. *Calibration using real observations.* We are now ready to calibrate the PSU3D-ICE model using the real observational data sets described in Section 2. Our main goal is to investigate whether using information on past grounding line position in our calibration leads to reduced uncertainty and better constrained projections. To this end, we calibrate the PSU3D-ICE model based on (i) only the modern binary patterns and (ii) both information sources. For (i) we conduct calibration using only the part of the posterior density that is related to \mathbf{Y}_2 and \mathbf{Z}_2 , and for (ii) we use the entire posterior density. We have obtained MCMC samples with 100,000 for (i) and 47,000 iterations for (ii), and checked its convergence and standard errors, as above.

The results clearly show the utility of using past grounding line positions in calibrating the PSD3D-ICE model. Figure 6 shows that using the past grounding line positions in our calibration drastically reduces parametric uncertainty. The ice volume change projection based on both information sources also has significantly less uncertainty than that based only on the modern binary patterns (Figure 7). In



(a) Modern binary patterns only

FIG. 6. Posterior density for input parameters based on the actual observational data sets (for details see Section 4.2). Using both information sources leads to significantly less uncertainty in estimating input parameters.

particular, using the grounding line positions in calibration eliminates the probability mass below 0 m sea level equivalent in the predictive distribution. This improvement is due to the fact that some parameter settings result in very unrealistic ice sheet behavior from the last glacial maximum to the present day, but give modern ice coverage patterns that are close to the current state. See the next subsection for further discussion.

Note that the calibration results in Figure 6(a) are somewhat different from the results based on the same observational data set shown by Chang et al. (2016) because there are 6 other parameters that are varied in the ensemble used by Chang et al. (2016), which are fixed by expert judgment in this experiment.

4.3. *Insights.* The results presented in the previous subsection clearly indicate that using the past grounding line position leads to better calibration results with less parametric uncertainties and, in turn, sharper future ice volume change pro-

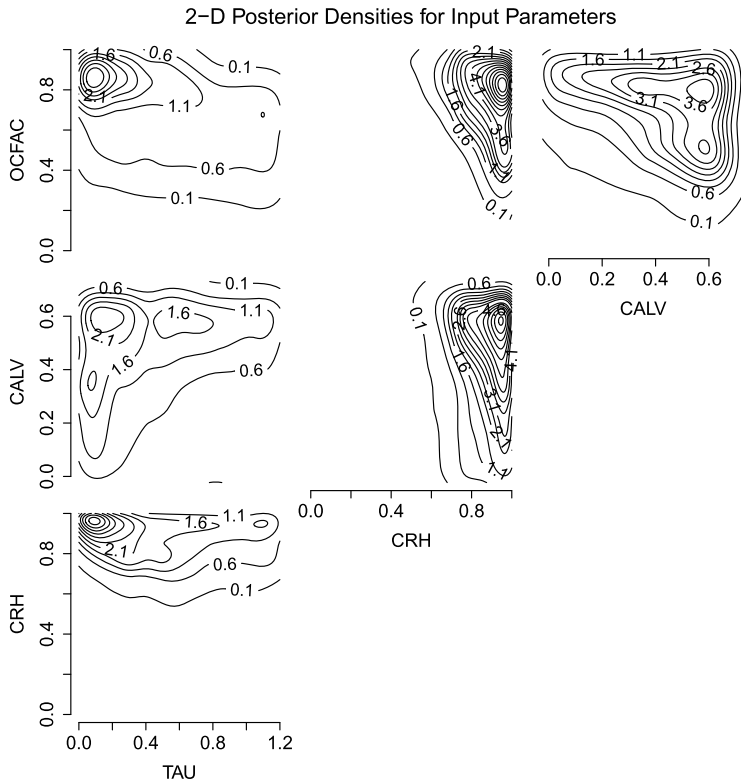


FIG. 6. (Continued).

jections. The hindcast and forecast of ice volume changes based on different data sources in Figure 8 clearly show the reason for this improvement. The 95% prediction intervals show that using the information from past grounding line positions significantly reduces the uncertainties in ice volume change trajectories by ruling out the parameter settings that generate unrealistic past ice sheet behavior. In particular, some parameter settings produce ice sheets that start from a very high ice volume around 15,000 years ago and then show unrealistically rapid ice volume loss, thereby resulting in a modern ice volume that is close to the observed value. These parameter settings are the cause of the left tail of volume change distributions based on the modern binary pattern only. Using paleo data rules out these parameter settings, thereby cutting off the left tail and reducing parametric uncertainty.

The credibility of this result of course depends on (i) how the model can reliably reproduce the past grounding line time series and (ii) how our calibration method accounts for the model-observation discrepancy. Figure 9 shows the grounding line position time series from the model outputs at the parameter settings $\theta_1, \dots, \theta_q$

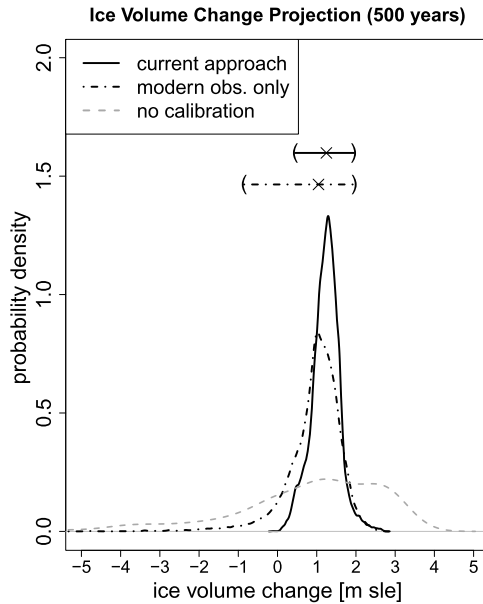


FIG. 7. Predictive distribution for the ice volume change projection based on the real observations described in Section 4.2. The black solid line shows the predictive distribution based on our approach using both the past grounding line positions and the modern binary patterns, and the black dashed and dotted line represents the result based on only the modern binary patterns. The gray dashed line shows the projection without calibration. The results show that using the past grounding line leads to a significantly sharper projection by removing the unrealistic ice volume increase in the results solely based on the modern observations.

and the observational data. It also shows the observational time series corrected by the discrepancy term $\mathbf{K}_{1,d}\mathbf{v}_1$. The discrepancy automatically accounts for the fact that all the model time series start further inland than the observed starting points. Hence, this is analogous to working with anomalies (difference between an observed value and a standard reference value). The use of anomalies is common in paleoclimate modeling; that is, the modeled *change* in a quantity is considered more reliable than the modeled *absolute* value because the model errors that remain constant through time cancel when differences are taken. The model change can either be the difference from the initial model state or the difference from a “control” simulation of an observed state. The difference is added to the observed quantity to yield a more robust model projection. This is equivalent here to uniformly shifting the observed grounding line positions to coincide with the mean model initial position at 15,000 years before present (Figure 9).

In addition, we also show another corrected observed time series in which only the starting position is matched with the discrepancy-adjusted observed time series discussed above. The difference between this time series (dashed-dotted line in Figure 9) and the fully corrected time series (dashed line in Figure 9) can be

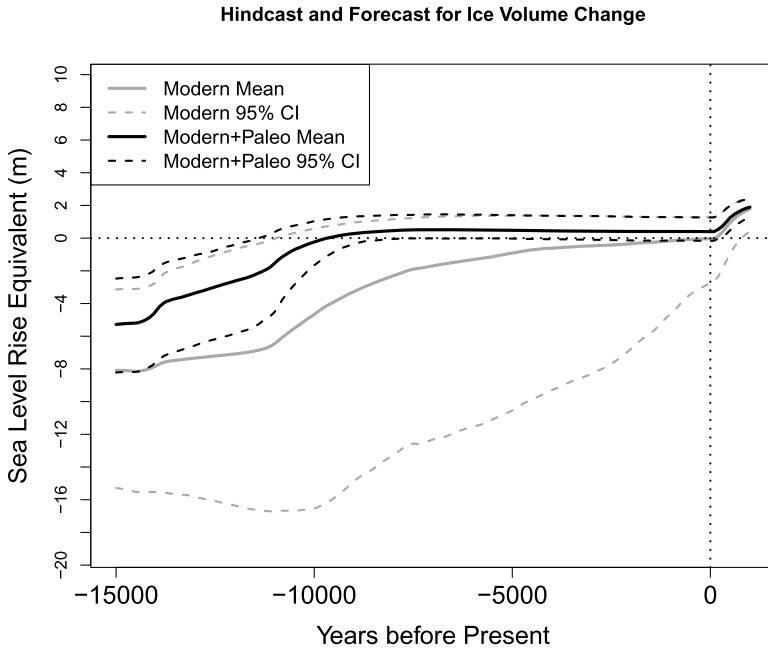


FIG. 8. The mean (solid lines) and pointwise 95% prediction limits (dashed lines) for projected ice volume changes based on the modern binary patterns only (gray) and the past grounding line positions and modern binary patterns (black). The negative values on the y-axis indicate the ice volume is larger than the modern value. The prediction limits based only on modern binary patterns contain trajectories that start from excessive amount of ice volume and show very fast ice volume decay. The prediction limits based on both sources of information rule out such unrealistic trajectories.

viewed as the estimated discrepancy in the anomaly. Although the modeled and observed anomalies clearly show a certain degree of discrepancy, the observed anomaly still allows us to rule out the model runs that showed too small or too large total grounding line position changes over 15,000 years. Therefore, the observed grounding line positions still provide useful information for reducing parametric uncertainty.

5. Discussion and caveats.

5.1. *Discussion.* In this work we have proposed a computationally efficient approach to calibrating WAIS models using two different sources of information, the past grounding line positions and the modern binary spatial patterns. Using the proposed approach, we have successfully calibrated the PSU-3D ice model and generated the WAIS volume change projections. Results from a simulated example indicate that our method recovers the true parameters with reasonably small uncertainties as well as provides useful information on interactions between parameters.

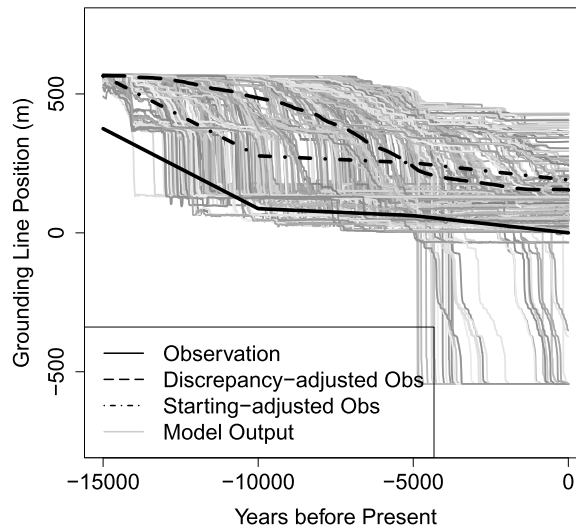


FIG. 9. The grounding line position time series from the observational data set (black solid line), observational data adjusted by the discrepancy term (black dashed line), observational data shifted to have the same starting point as the fully adjusted data (dashed and dotted black line), and the model outputs at the input parameter settings $\theta_1, \dots, \theta_p$ (gray lines). The discrepancy term shifts the observational data to match the starting grounding line position to the model outputs, which is similar to using anomalies.

Results based on the real observations indicate that using the paleo-record significantly reduces parametric uncertainty and leads to better constrained projections by removing the probability for unrealistic ice volume increase in the predictive distribution.

Several recent modeling studies of Antarctic Ice Sheet variations have used heuristic approaches to study parametric uncertainties, mostly applied to ice sheet retreat since the last glacial maximum about 15,000 years ago [Briggs, Pollard and Tarasov (2013, 2014), Briggs and Tarasov (2013), Golledge et al. (2014), Maris et al. (2015), Whitehouse, Bentley and Le Brocq (2012), Whitehouse et al. (2012)]. Using highly aggregated data and less statistically formal frameworks, these studies try to reduce the parametric uncertainties based on geologic data around Antarctica. By and large, our results are consistent with the parameter values found in these studies, while also reducing uncertainties about the parameters. Other recent modeling studies have projected future Antarctic Ice Sheet response to anthropogenic warming in coming centuries to millennia [Cornford et al. (2015), Feldmann and Levermann (2015), Golledge et al. (2015), Gomez, Pollard and Holland (2015), Ritz et al. (2015), Winkelmann et al. (2015)]. These models have been calibrated only over observed small-scale variations of the last few decades. Only a few of these studies use large ensembles, and none use the advanced statistical methods we have developed here, which allow for analyses based on large ensem-

bles and unaggregated data sets. Furthermore, we are able to obtain projections as well as parameter inference in the form of genuine probability distributions, and we take into account potential data-model discrepancies that are ignored by other studies. This allows us to provide uncertainties about our estimates and projections.

5.2. *Caveats and future directions.* One caveat in our calibration model specification is that we do not take into account the dependence between the past grounding line positions and the modern binary patterns. However, we note that the past grounding line positions and the modern binary patterns contain quite different information since two model runs with very different trajectories of past grounding line position often end up with very similar modern binary patterns; this is corroborated by an examination of cross-correlations. Developing a calibration approach based on the generalized principal component analysis that reduces the dimensionality of Gaussian and binary data simultaneously and computes common principal component scores for both data sets is one possible future direction.

Our results are also subject to the usual caveats in ice sheet modeling. For example, we use simplified atmospheric conditions for projections, which assume that atmospheric and oceanic temperatures linearly increase until 150 years after present and stay constant thereafter. Using more detailed warming scenarios is a subject for further work. Another caveat for the present study is the use of coarse-grid global ocean model results to parameterize past basal melting under floating ice shelves. Fine-grid modeling of ocean circulation in Antarctic embayments is challenging and a topic for further work [e.g., Hellmer et al. (2012)]. Another improvement will be the use of finer scale models with higher-order ice dynamics, which, as discussed in the Introduction, are not quite feasible for the large space and time scales of this study, but should gradually become practical in the near future.

APPENDIX A: COMPUTATION IN REDUCED-DIMENSIONAL SPACE

For faster computation we infer θ and other parameters in the model based on the following dimension-reduced version of the observational data for the past grounding line positions:

$$\mathbf{Z}_1^R = (\mathbf{K}_1^T \mathbf{K}_1)^{-1} \mathbf{K}_1^T \mathbf{Z}_1 = \begin{pmatrix} \eta(\theta, \mathbf{Y}_1^R) \\ \mathbf{v}_1 \end{pmatrix} + (\mathbf{K}_1^T \mathbf{K}_1)^{-1} \mathbf{K}_1^T \boldsymbol{\varepsilon}_1,$$

where $\mathbf{K}_1 = (\mathbf{K}_{1,y} \ \mathbf{K}_{1,d})$, which leads to the probability model

$$(8) \quad \mathbf{Z}_1^R | \mathbf{v}_2 \sim N \left(\begin{pmatrix} \boldsymbol{\mu}_\eta \\ \boldsymbol{\mu}_{\mathbf{v}_1 | \mathbf{v}_2} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_\eta & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{v}_1 | \mathbf{v}_2} \end{pmatrix} + \sigma_\varepsilon^2 (\mathbf{K}_1^T \mathbf{K}_1)^{-1} \right).$$

The J_1 -dimensional vector $\boldsymbol{\mu}_\eta$ and $J_1 \times J_1$ matrix $\boldsymbol{\Sigma}_\eta$ are the mean and variance of $\eta(\theta, \mathbf{Y}_1^R)$. The M -dimensional vector $\boldsymbol{\mu}_{\mathbf{v}_1 | \mathbf{v}_2}$ and $M \times M$ matrix $\boldsymbol{\Sigma}_{\mathbf{v}_1 | \mathbf{v}_2}$ are the

conditional mean and variance of \mathbf{v}_1 given \mathbf{v}_2 , which can be computed as

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{v}_1|\mathbf{v}_2} &= \frac{1}{\alpha_2^2} \mathbf{R}_v \mathbf{v}_2, \\ \boldsymbol{\Sigma}_{\mathbf{v}_1|\mathbf{v}_2} &= \alpha_1^2 (\mathbf{I}_M - \mathbf{R}_v \mathbf{R}_v^T). \end{aligned}$$

Using the likelihood function corresponding to this probability model and some standard prior specification for $\boldsymbol{\theta}$, α_1^2 and σ_ε^2 (see Section 4 for details), we can infer the parameters via Markov chain Monte Carlo (MCMC). The computational cost for likelihood evaluation reduces from $\frac{1}{3}n^3$ to $\frac{1}{3}(J_1 + M)^3$.

APPENDIX B: DETAILED DESCRIPTION FOR THE POSTERIOR DENSITY BASED ON THE MODEL SPECIFICATION IN SECTION 3.2.3

The parameters that we estimate in the equations in (5) and (6) are the input parameter $\boldsymbol{\theta}$ (which is our main target), the variance of the i.i.d. observational errors for the grounding line positions σ_ε^2 , coefficients for the emulator term $\boldsymbol{\psi} = \boldsymbol{\psi}(\boldsymbol{\theta}, \mathbf{Y}_2^R)$, the coefficients for the discrepancy term for the modern binary pattern \mathbf{v}_2 , and the variances of \mathbf{v}_1 and \mathbf{v}_2 , α_1^2 and α_2^2 . In addition to these parameters, we also re-estimate the sill parameters for the emulator $\boldsymbol{\eta}$, $\boldsymbol{\kappa}_1 = [\kappa_{1,1}, \dots, \kappa_{1,J_1}]$ [cf. Bayarri et al. (2007), Bhat et al. (2012), Chang et al. (2014a)] to account for a possible scale mismatch between the computer model output \mathbf{Y}_1 and the observational data \mathbf{Z}_1 . However, we do not re-estimate the sill parameters for $\boldsymbol{\psi}$, $\kappa_{2,1}, \dots, \kappa_{2,J_2}$, since both \mathbf{Y}_2 and \mathbf{Z}_2 are binary responses, and hence a scaling issue is not likely to occur here; in fact, we have found that re-estimating these parameters causes identifiability issues between the emulator term $\mathbf{K}_{2,y} \boldsymbol{\psi}(\boldsymbol{\theta}, \mathbf{Y}_2^R)$ and the discrepancy term $\mathbf{K}_{2,d} \mathbf{v}_2$.

The posterior density can be written as

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\kappa}_1, \mathbf{v}_2, \alpha_1^2, \alpha_2^2, \sigma_\varepsilon^2, \mathbf{R}_v | \mathbf{Y}_1^R, \mathbf{Z}_1^R, \mathbf{Y}_2^R, \mathbf{Z}_2) \\ \propto L(\mathbf{Z}_1^R | \mathbf{Y}_1^R, \boldsymbol{\theta}, \boldsymbol{\kappa}_1, \alpha_1^2, \sigma_\varepsilon^2, \mathbf{v}_2, \mathbf{R}_v) \\ \times f(\boldsymbol{\kappa}_1) f(\alpha_1^2) f(\sigma_\varepsilon^2) f(\mathbf{R}_v) \\ \times L(\mathbf{Z}_2 | \boldsymbol{\psi}, \mathbf{v}_2) \\ \times f(\boldsymbol{\psi} | \boldsymbol{\theta}, \mathbf{Y}_2^R) f(\mathbf{v}_2 | \alpha_2^2) f(\alpha_2^2) \\ \times f(\boldsymbol{\theta}). \end{aligned}$$

The likelihood function $L(\mathbf{Z}_1^R | \mathbf{Y}_1^R, \boldsymbol{\theta}, \boldsymbol{\kappa}_1, \alpha_1^2, \sigma_\varepsilon^2, \mathbf{v}_2, \mathbf{R}_v)$ is given by the probability model in (8). For $f(\boldsymbol{\kappa}_1) = f(\kappa_{1,1}, \dots, \kappa_{1,J_1})$ we use independent inverse gamma priors with a shape parameter of 50 and scale parameters specified in a way that the modes of the densities coincide with the estimated values of $\kappa_{1,1}, \dots, \kappa_{1,J_1}$ from the emulation stage. We assign a vague prior $IG(2, 3)$ for $f(\alpha_1^2)$, $f(\alpha_2^2)$ and

$f(\sigma_\varepsilon^2)$, and a uniform prior for $f(\boldsymbol{\theta})$ whose support is defined by the range of design points $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p$. The likelihood function $L(\mathbf{Z}_2 | \boldsymbol{\psi}, \mathbf{v}_2)$ is defined as

$$L(\mathbf{Z}_2 | \boldsymbol{\psi}, \mathbf{v}_2) \propto \prod_{j=1}^n \left(\frac{\exp(\lambda_j)}{1 + \exp(\lambda_j)} \right)^{Z_2(s_j)} \left(\frac{1}{1 + \exp(\lambda_j)} \right)^{1 - Z_2(s_j)},$$

where λ_j is the j th element of $\boldsymbol{\lambda}$ in (6). The conditional density $f(\boldsymbol{\psi} | \boldsymbol{\theta}, \mathbf{Y}_2^R)$ is given by the Gaussian process emulator $\boldsymbol{\psi}(\boldsymbol{\theta}, \mathbf{Y}_2^R)$. The conditional density $f(\mathbf{v}_2 | \alpha_2^2)$ is defined by the model $\mathbf{v}_2 \sim N(0, \alpha_2^2 \mathbf{I}_L)$. The prior density $f(\mathbf{R}_v)$ is defined as

$$\prod_{i=1}^M \prod_{j=1}^L I(-1 < \rho_{v,i,j} < 1) \cdot I(\mathbf{I}_M - \mathbf{R}_v \mathbf{R}_v^T \text{ is positive definite}),$$

where $I(\cdot)$ is the indicator function and $\rho_{v,i,j}$ is the (i, j) th element of \mathbf{R}_v .

Acknowledgments. We are grateful to A. Landgraf for distributing his code for logistic PCA freely on the Web (<https://github.com/andland/SparseLogisticPCA>). All views, errors and opinions are solely those of the authors.

SUPPLEMENTARY MATERIAL

Supplement to “Improving ice sheet model calibration using paleo and 2 modern observations: A reduced dimensional approach” (DOI: [10.1214/16-AOAS979SUPP](https://doi.org/10.1214/16-AOAS979SUPP); .pdf). We provide additional supporting plots that show more example model outputs for modern binary patterns and the leading principal components used in our calibration method.

REFERENCES

- APPLEGATE, P. J., KIRCHNER, N., STONE, E. J., KELLER, K. and GREVE, R. (2012). An assessment of key model parametric uncertainties in projections of Greenland ice sheet behavior. *Cryosphere* **6** 589–606.
- BAYARRI, M. J., BERGER, J. O., CAPEO, J., GARCIA-DONATO, G., LIU, F., PALOMO, J., PARTHASARATHY, R. J., PAULO, R., SACKS, J. and WALSH, D. (2007). Computer model validation with functional output. *Ann. Statist.* **35** 1874–1906. [MR2363956](#)
- BHAT, K. S., HARAN, M. and GOES, M. (2010). Computer model calibration with multivariate spatial output: A case study. In *Frontiers of Statistical Decision Making and Bayesian Analysis* (M. H. Chen, P. Müller, D. Sun, K. Ye and D. K. Dey, eds.) 168–184. Springer, New York.
- BHAT, K. S., HARAN, M., OLSON, R. and KELLER, K. (2012). Inferring likelihoods and climate system characteristics from climate models and multiple tracers. *Environmetrics* **23** 345–362. [MR2935569](#)
- BINDSCHADLER, R. A., NOWICKI, S., ABE-OUCHI, A., ASCHWANDEN, A., CHOI, H., FASTOOK, J., GRANZOW, G., GREVE, R., GUTOWSKI, G., HERZFELD, U., JACKSON, C., JOHNSON, J., KHROULEV, C., LEVERMANN, A., LIPSCOMB, W. H., MARTIN, M. A., MORLIGHEM, M., PARIZEK, B. R., POLLARD, D., PRICE, S. F., REN, D., SAITO, F., SATO, T., SEDDIK, H., SEROUSSI, H., TAKAHASHI, K., WALKER, R. and WANG, W. L. (2013). Ice-sheet model sensitivities to environmental forcing and their use in projecting future sea level (the SeaRISE project). *J. Glaciol.* **59** 195–224.

- BRIGGS, R., POLLARD, D. and TARASOV, L. (2013). A glacial systems model configured for large ensemble analysis of Antarctic deglaciation. *Cryosphere* **7** 1533–1589.
- BRIGGS, R. D., POLLARD, D. and TARASOV, L. (2014). A data-constrained large ensemble analysis of Antarctic evolution since the Eemian. *Quat. Sci. Rev.* **103** 91–115.
- BRIGGS, R. D. and TARASOV, L. (2013). How to evaluate model-derived deglaciation chronologies: A case study using Antarctica. *Quat. Sci. Rev.* **63** 109–127.
- BRYNJARSDÓTTIR, J. and O’HAGAN, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Probl.* **30** 114007, 24. [MR3274591](#)
- CHANG, W., HARAN, M., OLSON, R. and KELLER, K. (2014a). Fast dimension-reduced climate model calibration and the effect of data aggregation. *Ann. Appl. Stat.* **8** 649–673. [MR3262529](#)
- CHANG, W., APPEGATE, P., HARAN, H. and KELLER, K. (2014b). Probabilistic calibration of a Greenland ice sheet model using spatially-resolved synthetic observations: Toward projections of ice mass loss with uncertainties. *Geosci. Model Dev.* **7** 1933–1943.
- CHANG, W., HARAN, M., APPEGATE, P. and POLLARD, D. (2016). Supplement to “Improving ice sheet model calibration using paleoclimate and modern data.” DOI:[10.1214/16-AOAS979SUPP](#).
- CHANG, W., HARAN, M., APPEGATE, P. and POLLARD, D. (2016). Calibrating an ice sheet model using high-dimensional binary spatial data. *J. Amer. Statist. Assoc.* **111** 57–72. [MR3494638](#)
- CORNFORD, S. L., MARTIN, D. F., PAYNE, A. J., NG, E. G., LE BROCCQ, A. M., GLADSTONE, R. M., EDWARDS, T. L., SHANNON, S. R., AGOSTA, C., VAN DEN BROEKE, M. R., HELLMER, H. H., KRINNER, G., LIGTENBERG, S. R. M., TIMMERMANN, R. and VAUGHAN, D. G. (2015). Century-scale simulations of the response of the West Antarctic Ice Sheet to a warming climate. *Cryosphere* **9** 1579–1600.
- FAVIER, L., DURAND, G., CORNFORD, S. L., GUDMUNDSSON, G. H., GAGLIARDINI, O., GILLET-CHAULET, F., ZWINGER, T., PAYNE, A. J. and LE BROCCQ, A. M. (2014). Retreat of Pine Island Glacier controlled by marine ice-sheet instability. *Nature Climate Change* **4** 171–121.
- FELDMANN, J. and LEVERMANN, A. (2015). Collapse of the West Antarctic Ice Sheet after local destabilization of the Amundsen Basin. *Proc. Natl. Acad. Sci. USA* **112** 14191–14196.
- FLEGAL, J. M., HARAN, M. and JONES, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist. Sci.* **23** 250–260. [MR2516823](#)
- FRETWELL, P., PRITCHARD, H. D., VAUGHAN, D. G., BAMBER, J. L., BARRAND, N. E., BELL, R., BIANCHI, C., BINGHAM, R. G., BLANKENSHIP, D. D., CASASSA, G., CATANIA, G., CALLENS, D., CONWAY, H., COOK, A. J., CORR, H. F. J., DAMASKE, D., DAMM, V., FERRACCIOLI, F., FORSBERG, R., FUJITA, S., GIM, Y., GOGINENI, P., GRIGGS, J. A., HINDMARSH, R. C. A., HOLMLUND, P., HOLT, J. W., JACOBEL, R. W., JENKINS, A., JOKAT, W., JORDAN, T., KING, E. C., KOHLER, J., KRABILL, W., RIGER-KUSK, M., LANGLEY, K. A., LEITCHENKOV, G., LEUSCHEN, C., LUYENDYK, B. P., MATSUOKA, K., MOUGINOT, J., NITSCHKE, F. O., NOGI, Y., NOST, O. A., POPOV, S. V., RIGNOT, E., RIPPIN, D. M., RIVERA, A., ROBERTS, J., ROSS, N., SIEGERT, M. J., SMITH, A. M., STEINHAGE, D., STUDINGER, M., SUN, B., TINTO, B. K., WELCH, B. C., WILSON, D., YOUNG, D. A., XIANGBIN, C. and ZIRIZZOTTI, A. (2013). Bedmap2: Improved ice bed, surface and thickness datasets for Antarctica. *Cryosphere* **7** 375–393.
- GLADSTONE, R. M., LEE, V., ROUGIER, J., PAYNE, A. J., HELLMER, H., LE BROCCQ, A., SHEPHERD, A., EDWARDS, T. L., GREGORY, J. and CORNFORD, S. L. (2012). Calibrated prediction of Pine Island Glacier retreat during the 21st and 22nd centuries with a coupled flowline model. *Earth Planet. Sci. Lett.* **333** 191–199.
- GOLLEDGE, N. R., MENVIEL, L., CARTER, L., FOGWILL, C. J., ENGLAND, M. H., CORTESE, G. and LEVY, R. H. (2014). Antarctic contribution to meltwater pulse 1A from reduced Southern Ocean overturning. *Nature Comm.* **5**.
- GOLLEDGE, N. R., KOWALEWSKI, D. E., NAISH, T. R., LEVY, R. H., FOGWILL, C. J. and GASSON, E. G. W. (2015). The multi-millennial Antarctic commitment to future sea-level rise. *Nature* **526** 421–425.

- GOMEZ, N., POLLARD, D. and HOLLAND, D. (2015). Sea-level feedback lowers projections of future Antarctic ice-sheet mass loss. *Nature Comm.* **6**.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York. [MR2722294](#)
- HELLMER, H. H., KAUKER, F., TIMMERMANN, R., DETERMANN, J. and RAE, J. (2012). Twenty-first-century warming of a large Antarctic ice-shelf cavity by a redirected coastal current. *Nature* **485** 225–228.
- HIGDON, D., GATTIKER, J., WILLIAMS, B. and RIGHTLEY, M. (2008). Computer model calibration using high-dimensional output. *J. Amer. Statist. Assoc.* **103** 570–583. [MR2523994](#)
- JONES, G. L., HARAN, M., CAFFO, B. S. and NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547. [MR2279478](#)
- JOUGHIN, I., SMITH, B. E. and MEDLEY, B. (2014). Marine ice sheet collapse potentially under way for the Thwaites Glacier Basin, West Antarctica. *Science* **344** 735–738.
- KENNEDY, M. C. and O’HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 425–464. [MR1858398](#)
- KIRSHNER, A. E., ANDERSON, J. B., JAKOBSSON, M., O’REGAN, M., MAJEWSKI, W. and NITSCHKE, F. O. (2012). Post-LGM deglaciation in Pine Island Bay, West Antarctica. *Quat. Sci. Rev.* **38** 11–26.
- LARTER, R. D., ANDERSON, J. B., GRAHAM, A. G., GOHL, K., HILLENBRAND, C.-D., JAKOBSSON, M., JOHNSON, J. S., KUHN, G., NITSCHKE, F. O. and SMITH, J. A. (2014). Reconstruction of changes in the Amundsen Sea and Bellingshausen sea sector of the West Antarctic Ice Sheet since the last glacial maximum. *Quat. Sci. Rev.* **100** 55–86.
- LEMPERT, R., SRIVER, R. L. and KELLER, K. (2012). *Characterizing uncertain sea level rise projections to support investment decisions*. California Energy Commission. Publication Number: CEC-500-2012-056.
- LIU, Z., OTTO-BLIESNER, B. L., HE, F., BRADY, E. C., TOMAS, R., CLARK, P. U., CARLSON, A. E., LYNCH-STIEGLITZ, J., CURRY, W., BROOK, E., ERICKSON, D., JACOB, R., KUTZBACH, J. and CHENG, J. (2009). Transient simulation of last deglaciation with a new mechanism for Bølling–Allerød warming. *Science* **325** 310–314.
- MARIS, M. N. A., VAN WESSEME, J. M., VAN DE BERG, W. J., DE BOER, B. and OERLEMANS, J. (2015). A model study of the effect of climate and sea-level change on the evolution of the Antarctic Ice Sheet from the Last Glacial Maximum to 2100. *Clim. Dynam.* **45** 837–851.
- MCNEALL, D. J., CHALLENGER, P. G., GATTIKER, J. R. and STONE, E. J. (2013). The potential of an observational data set for calibration of a computationally expensive computer model. *Geosci. Model Dev.* **6** 1715–1728.
- POLLARD, D. and DECONTO, R. M. (2009). Modelling West Antarctic Ice Sheet growth and collapse through the past five million years. *Nature* **458** 329–332.
- POLLARD, D. and DECONTO, R. M. (2012a). A simple inverse method for the distribution of basal sliding coefficients under ice sheets, applied to Antarctica. *Cryosphere* **6** 1405–1444.
- POLLARD, D. and DECONTO, R. M. (2012b). Description of a hybrid ice sheet-shelf model, and application to Antarctica. *Geosci. Model Dev.* **5** 1273–1295.
- PRITCHARD, H. D., LIGTENBERG, S. R. M., FRICKER, H. A., VAUGHAN, D. G., VAN DEN BROEKE, M. R. and PADMAN, L. (2012). Antarctic ice-sheet loss driven by basal melting of ice shelves. *Nature* **484** 502–505.
- RAISED CONSORTIUM (2014). A community-based geological reconstruction of Antarctic Ice Sheet deglaciation since the Last Glacial Maximum. *Quat. Sci. Rev.* **100** 1–9.
- RITZ, C., EDWARDS, T. L., DURAND, G., PAYNE, A. J., PEYAUD, V. and HINDMARSH, R. C. (2015). Potential sea-level rise from Antarctic ice-sheet instability constrained by observations. *Nature* **528** 115–118.
- SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4** 409–435. [MR1041765](#)

- STONE, E. J., LUNT, D. J., RUTT, I. C. and HANNA, E. (2010). Investigating the sensitivity of numerical model simulations of the modern state of the Greenland ice-sheet and its future response to climate change. *Cryosphere* **4** 397–417.
- WHITEHOUSE, P. L., BENTLEY, M. J. and LE BROCCQ, A. M. (2012). A deglacial model for Antarctica: Geological constraints and glaciological modeling as a basis for a new model of Antarctic glacial isostatic adjustment. *Quat. Sci. Rev.* **32** 1–24.
- WHITEHOUSE, P. L., BENTLEY, M. J., MILNE, G. A., KING, M. A. and THOMAS, I. D. (2012). A new glacial isostatic model for Antarctica: Calibrated and tested using observations of relative sea-level change and present-day uplifts. *Geophysical Journal International* **190** 1464–1482.
- WINKELMANN, R., LEVERMANN, A., RIDGWELL, A. and CALDEIRA, K. (2015). Combustion of available fossil fuel resources sufficient to eliminate the Antarctic Ice Sheet. *Sci. Adv.* **1** e1500589.

W. CHANG
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF CINCINNATI
CINCINNATI, OHIO 45221
USA
E-MAIL: won.chang@uc.edu

M. HARAN
DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802
USA
E-MAIL: mharan@stat.psu.edu

P. APPLGATE
D. POLLARD
EARTH AND ENVIRONMENTAL SYSTEMS INSTITUTE
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802
USA
E-MAIL: patrick.applegate@psu.edu
pollard@essc.psu.edu