ICE MODEL CALIBRATION USING SEMI-CONTINUOUS SPATIAL DATA

BY WON CHANG*, BLEDAR A. KONOMI*, GEORGIOS KARAGIANNIS†, YAWEN GUAN‡ AND MURALI HARAN§

University of Cincinnati*, Durham University†, University of Nebraska‡, The Pennsylvania State University§

Rapid changes in Earth’s cryosphere caused by human activity can lead to significant environmental impacts. Computer models provide a useful tool for understanding the behavior and projecting the future of Arctic and Antarctic ice sheets. However, these models are typically subject to large parametric uncertainties due to poorly constrained model input parameters that govern the behavior of simulated ice sheets. Computer model calibration provides a formal statistical framework to infer parameters using observational data, and to quantify the uncertainty in projections due to the uncertainty in these parameters. Calibration of ice sheet models is often challenging because the relevant model output and observational data take the form of semi-continuous spatial data, with a point mass at zero and a right-skewed continuous distribution for positive values. Current calibration approaches cannot handle such data. Here we introduce a hierarchical latent variable model that handles binary spatial patterns and positive continuous spatial patterns as separate components. To overcome challenges due to high-dimensionality we use likelihood-based generalized principal component analysis to impose low-dimensional structures on the latent variables for spatial dependence. We apply our methodology to calibrate a physical model for the Antarctic ice sheet and demonstrate that we can overcome the aforementioned modeling and computational challenges. As a result of our calibration, we obtain improved future ice-volume change projections.

1. Introduction. Human-induced climate change is projected to significantly affect the Earth’s cryosphere. The West Antarctic ice sheet (WAIS) is particularly susceptible to warming climate because a large portion of its body is marine based, meaning that the bottom of the ice is below the sea-level. Any significant changes in this part of Antarctica can lead to a consequential sea level change (Fretwell et al., 2013). Computer models are used to project the future of WAIS, but the projections from these computer models are highly uncertain due to uncertainty about the values of key model input parameters (Stone et al., 2010; Gladstone et al., 2012; Chang et al., 2016a; Pollard et al., 2016). Computer model calibration provides a statistical framework for using observational data to infer input parameters of complex computer models.

Following the calibration framework described in the seminal paper by Kennedy and O’Hagan (2001), several researchers have developed methods for inferring model parameters for a variety of different types of computer model output. For instance, Bayarri et al. (2007) provides a wavelet-based approach for calibration with functional model output. Sansó and Forest (2009) calibrates a climate model with multivariate output while Higdon et al. (2008)

Keywords and phrases: Computer Model Emulation, Computer Model Calibration, Semi-continuous Spatial Data, West Antarctic Ice Sheet, Dimension Reduction
and Chang et al. (2014) provide approaches for calibrating models with high-dimensional spatial data output. More recently, Chang et al. (2016a) develops an approach for high-dimensional binary spatial data output and Sung et al. (2019) proposes a method for binary time series output. Cao et al. (2018) provides a method for censored functional data. Ice sheet thickness data, including the West Antarctic ice sheet data set we analyse here, are frequently in the form of high-dimensional semi-continuous spatial data. No existing calibration methods are suited to this type of data; this motivates the new methodological development in this manuscript.

Several computer model calibration approaches have been applied to infer the parameters and to systematically quantify parametric uncertainty in Antarctic ice sheet models (Gladstone et al., 2012; Chang et al., 2016a,b; Pollard et al., 2016; Edwards et al., 2019). One important caveat to existing approaches to ice sheet model calibration is that the model outputs and observational data need to be transformed or aggregated in some degree to avoid issues involving semi-continuous distributions. To be more specific, the main variable of interest in ice model output and observational data is the spatial pattern of ice thicknesses which takes positive values at locations with ice presence and zero values otherwise. Handling such spatially dependent semi-continuous data with truncation at zero poses non-trivial inferential and computational challenges and existing calibration methods cannot readily handle these issues. Chang et al. (2016a) used ice-no ice binary spatial patterns obtained by dichotomizing the thickness patterns into zeros and ones and hence ignored important information regarding the ice thickness. Pollard et al. (2016) also similarly used highly-summarized data to avoid challenges related to semi-continuous data. Although their results show that such approaches still lead to a meaningful reduction in input parameter uncertainty, one can certainly expect that transforming or summarizing data can result in a significant loss of information. This motivates our methodological development of a calibration method that can directly utilize semi-continuous spatial data.

Existing methods for handling semi-continuous data in the spatial statistics literature are based on the truncated Gaussian process approach (Stein, 1992; De Oliveira, 2005). In this framework the semi-continuous data being analyzed are viewed as a realization from an underlying Gaussian process (GP), which can be observed only when the values are positive. This simple ‘clipped’ Gaussian process approach provides a natural way to impose spatial dependence among zero and non-zero values. However, using a truncated Gaussian process can create serious computational issues when applied to a high dimensional data set with a large proportion of zeros. This is because inference based on such a model requires integrating out highly-dependent, high-dimensional, and bounded latent variables for locations with zero values. Matrix computations for high-dimensional spatial random variables are expensive. Furthermore, designing efficient (‘fast mixing’) Markov chain Monte Carlo methods for Bayesian inference for such models becomes very challenging. This is why a clipped Gaussian process (such as the one used by Cao et al., 2018) cannot provide a feasible solution for our calibration problem.

In this paper we formulate an emulation and calibration framework that uses two separate processes: one process for modeling the presence and absence of ice and the other for modeling the value of ice thickness given that ice is present. This approach removes the need to integrate out the bounded latent variables for the locations with no ice and hence allows us to circumvent the related computational challenges in the clipped Gaussian process ap-
ICE MODEL CALIBRATION USING SEMI-CONTINUOUS SPATIAL DATA

Our proposed method uses likelihood-based principal component analysis (Tipping and Bishop, 1999) to reduce the dimension of model output and observational data (cf. Higdon et al., 2008; Chang et al., 2014), and avoids issues with large non-Gaussian spatial data calibration (cf. Chang et al., 2016a). In our simulated example and real data analysis, we show that our method can efficiently utilize information from large semi-continuous spatial data and lead to improved calibration results compared to using only binary spatial patterns. While our focus is on calibrating a computer model for WAIS, the methodology we develop here is readily applicable, with only minor modifications, to other calibration problems with semi-continuous data.

The rest of this paper is organized as follows. In Section 2, we introduce the details of our PSU-3D model runs and Bedmap2 observational data that have motivated our methodological development. In Section 3, we describe our new framework for emulation and calibration using semi-continuous data and discuss the computational challenges posed by the large size of the spatial data. In Section 4, we propose a reduced-dimension approach that can mitigate the computational challenges, and in Section 5 we describe the result of our analysis on the model runs and observational data using the proposed approach. In Section 6 we summarize our findings and discuss some possible future directions.

2. Model Runs and Observational Data. In this study we use a state-of-the-art model, the PSU-3D ice model (Pollard et al., 2015, 2016), for studying the evolution of WAIS. This model strikes a good balance between model realism and computational efficiency and hence can allow simulations of long term behavior of WAIS (on the scale of thousands of years) with a relatively high resolution of 20 km. Similar to other complex computer model experiments, simulation runs from the PSU-3D ice model are available only at a limited number of input parameter settings due to the high computational cost. Therefore in this study we take an emulation approach in which we first create a collection of model runs at pre-specified design points in the input parameter space (often called a perturbed physics ensemble) and then build a statistical surrogate based on those model runs.

We use a previously published ensemble of simulations (Chang et al., 2016a) generated from PSU-3D ice model with 499 model runs. The parameter settings for ensemble members are determined by a Latin hypercube design for ten varied input parameters. Note that a part of the simulation results were analyzed by Chang et al. (2016a) in which the simulated modern ice thickness patterns are dichotomized into binary spatial patterns and used for calibration. In this study we aim to utilize the full ice thickness patterns for the modern WAIS as well as propose a statistical approach that can properly handle the inferential challenges arising from the data type of the thickness patterns, semi-continuous spatial data with many zeros.

We refer to Table S1 in the Supporting Information for detailed description about these ten parameters calibrated in this paper. (For the rest of the paper, the figure and table numbers starting with ‘S’ refer to the ones in the Supporting Information.) The ranges of these parameters are determined based on the previous literature (Pollard and DeConto, 2012; Pollard et al., 2015). While these parameters play important roles in determining the long-term evolution of the Antarctic ice sheet, their values are highly uncertain and hence need to be properly calibrated for realistic simulation.
Each ensemble member is spun up for 40,000 years to the present and then projected into future for 5,000 years. For ocean forcing, the model runs use archived ocean temperatures from a coupled AOGCM simulation by Liu et al. (2009). For atmospheric forcing, we use a modern temperature map (ALBMAP, Le Brocq, A.M. and Payne, A.J. and Vieli, A., 2010) that is uniformly perturbed based on a deep-sea-core δ18O record (Pollard and DeConto, 2009, 2012). For future projection we assume that the oceanic temperature linearly increases over the first 150 years to reach 2 °C of warming and then remains constant thereafter. The atmospheric temperature also increases in the same way to reach 2 °C of warming. Creating more realistic future forcing scenarios for Antarctic simulation is currently an active research area (see, e.g., Berdahl et al., 2020) and we leave it as a future work. Once the simulation over the entire WAIS is done we extract the spatial pattern of modern grounded ice sheet thickness in Amundsen Sea Embayment (ASE) region, which is expected to be one of the major contributors to sea level change in the future. The spatial pattern in our selected region has 86 × 37 pixels with 20 km × 20 km resolution (Figures 1 b-d). Figure 1e (adapted from Pollard et al., 2015) shows the location of the study region in WAIS and the topography of the bedrock in the area.

To calibrate the ten input parameters varied in the ensemble, we compare these model outputs with the observed modern ice sheet thickness pattern in the same area derived from the Bedmap2 dataset (Fretwell et al., 2013) (Figure 1a). This recent data product combines a wide range of sources including seismic sounding, radar surveys, and satellite altimetry. Since the observational grid has a higher spatial resolution (1 km × 1 km resolution), we upscale the observational data to the model grid by integrating over 400 observational grid cells for each model grid cell. Note that the model outputs and the observational data for ice thickness are all in the form of high-dimensional semi-continuous spatial data which poses non-trivial statistical challenges for our calibration framework.

3. Computer Model Emulation and Calibration Using Semi-continuous Spatial Data. In this section we describe our statistical framework for inferring the input parameters in the PSU-3D ice model. In particular we focus on describing how the standard computer model emulation and calibration framework (Kennedy and O’Hagan, 2001) can be modified to accommodate the ice thickness patterns introduced above, which take the form of semi-continuous data. A flowchart describing the overall modelling framework can be found in Figure S1.

We use the following notation hereafter: Let the p-dimensional vector \( Y(\theta) = [Y(\theta, s_1), \ldots, Y(\theta, s_p)]^T \) denote the spatial pattern of ice thickness at the spatial locations of the model grid \( s_1, \ldots, s_p \in \mathbb{R}^2 \) which is generated from the computer model given input parameter setting \( \theta \in \mathbb{R}^d \). Here, \( d \) is the dimension of the input space which in our application is equal to ten. The observed data at the same spatial locations are denoted as a p-dimensional vector \( Z = [Z(s_1), \ldots, Z(s_p)]^T \). Here, \( Y(\theta, s_j) \) and \( Z(s_j) \) can have either positive values representing the ice thickness or zero values denoting absence of ice at location \( s_j \) (see Figure 1).

We denote the design points for the input parameters in our ensemble as \( \theta_1, \ldots, \theta_n \). As a result the collection of model output in our ensemble can be denoted as an \( n \times p \) matrix \( Y \), with elements \( Y_{i,j} = Y(\theta_i, s_j) \) for \( i = 1, \ldots, n \) and \( j = 1, \ldots, p \), where the rows correspond to different input parameter settings while the columns correspond to different
Fig 1: Observational data (a) from Bedmap 2 data (Fretwell et al., 2013) and example model runs (b-d) from PSU-3D ice model. (e) The maps of region names (left) and bedrock elevations (right) that are adopted from Pollard et al. (2015), with a red box showing the study area. The ice sheet in the study area is mostly marine-based, i.e., its bottom is beneath the sea level.
spatial locations. In our ice thickness application the number of spatial locations for the
grid is $p = 86 \times 37 = 3,182$ and the number of model runs in the ensemble is $n = 499$.

3.1. Procedure Overview. Since our methodological development involves a lengthy dis-
\text{cussion, we first give a preview of the overall steps of our approach. Given the $n \times p$ matrix
for model output $Y$ and $p$-dimensional vector for observational data $Z$:}

1. Create a $n \times p$ matrix for ice-no ice binary patters, $\{I_y(\theta_i, s_j)\}$ ($i = 1, \ldots, n$, $j = 1, \ldots, p$), by dychotomizing the elements in the model output matrix $Y$ into 0s and
\text{1s. Apply logistic principal component analysis (LPCA) to the dychotomized output
matrix to find the $n \times J_w$ matrix for LPC scores $W$.}

2. \text{Apply likelihood-based PCA only to the non-zero values in $Y$, to find the $n \times J_u$
matrix for PC scores $U$.}

3. \text{For each column in $W$ and $U$, separately construct a 1-dimensional GP emulator by
finding Bayesian estimates for the emulator parameters with the procedure proposed
by Gu et al. (2019). Let $\psi(\theta)$ and $\xi(\theta)$ respectively denote $J_w$- and $J_u$-dimensional
emulated processes for the unobserved values of $w(\theta)$ and $u(\theta)$, which are collections
of independently constructed 1-dimensional GP emulators.}

4. Infer the best input parameter setting $\theta^*$ along with other parameters based on the
posterior density given the observational data $Z$ (see Equation (4.7) for its definition).
\text{The Bayesian inference can be facilitated through Metropolis-within-Gibbs sampling.}

A flowchart describing the overall modeling framework is included as Figure S1 in SI.

3.2. Computer Model Emulation Using Semi-Continuous Spatial Data. Since only a lim-
\text{ited number of computer runs can be carried out, we use an emulator to statistically link
the modeled ice thickness to the observational data. However, the semi-continuous nature
of $Y(\theta)$ prevents direct application of existing GP calibration approaches such as those
in Sacks et al. (1989) and Kennedy and O’Hagan (2001). In order to make emulation of
the semi-continuous $Y(\theta, s_j)$ variable possible, we introduce an indicator variable $I_y(\theta_i, s_j)
whose value is one if grounded ice is present at the given parameter setting and spatial loca-
tion $(\theta_i, s_j)$ or zero otherwise for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Given that grounded ice is
present, we model the thickness as $Y(\theta_i, s_j) = q(h(\theta_i, s_j))$, where $q : \mathbb{R} \rightarrow \mathbb{R}_+$ is a bijective
transformation function that allows $h(\theta_i, s_j)$ to take any real value. (See Section 5.3 for
our choice for $q$ and reasoning for it.) We can now formulate the ice thickness $Y(\theta_i, s_j)$ as

\begin{equation}
Y(\theta_i, s_j) = \begin{cases} 
q(h(\theta_i, s_j)) & \text{if } I_y(\theta_i, s_j) = 1 \\
0 & \text{if } I_y(\theta_i, s_j) = 0 \end{cases},
\end{equation}

for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Using this representation, we can translate the
problem of emulating $Y(\theta)$ into the problem of finding the predictive distributions of the binary
response $I_y(\theta, s_1), \ldots, I_y(\theta, s_p)$ and the transformed thickness values $h(\theta) = [h(\theta, s_1), \ldots,
h(\theta, s_p)]^T$ at any untried input parameter setting $\theta$. Therefore, we can model $h(\theta)$ directly
using an existing method for continuous data such as basis representation (see, e.g., Higdon
et al., 2008; Chang et al., 2014), since its elements are unbounded and continuous. We use
a $p$-dimensional vector $\eta(\theta) = [\eta(\theta, s_1), \ldots, \eta(\theta, s_p)]^T$ to denote the emulated process for
$h(\theta)$. (The actual emulation will be done using a basis representation method as described in}
Section 4.1 below). We indirectly emulate the binary spatial pattern \( I_y(\theta, s_1), \ldots, I_y(\theta, s_p) \) through their corresponding logits \( \gamma(\theta) = [\gamma(\theta, s_1), \ldots, \gamma(\theta, s_p)]^T \) defined as

\[
\gamma(\theta, s_j) = \log \frac{P(I_y(\theta, s_j) = 1)}{1 - P(I_y(\theta, s_j) = 1)}
\]

for \( j = 1, \ldots, p \) as in Chang et al. (2016a). Since \( \gamma(\theta) = [\gamma(\theta, s_1), \ldots, \gamma(\theta, s_p)]^T \) can be again treated as continuous variables with unbounded support, an emulation approach for continuous variables can be applicable. Since \( \gamma(\theta, s_j) \) is an unobserved latent variable even if \( \theta \) is one of the existing design points \( \theta_1, \ldots, \theta_n \), we do not use a separate notation for the logits at those design points. Our emulation problem now becomes a problem of finding predictive processes \( \eta(\theta) \) and \( \gamma(\theta) \) at any untried settings \( \theta \) (which are possibly dependent on each other).

3.3. Computer Model Calibration Using Semi-Continuous Spatial Data. In addition to the new emulation framework described above we formulate a new calibration framework for semi-continuous data because the standard calibration approach (Kennedy and O’Hagan, 2001) is not applicable. Here we use a similar representation of the observed ice thickness \( Z(s_j) \) as in (3.1). We define the variable \( I_z(s_j) \) to be an indicator with a value of one if observed grounded ice presents at \( s_j \) and zero otherwise. To transform the observational data, we use the same transformation function \( q \) as in (3.1). At any spatial location \( s_j \), we assume observation of ice thickness \( Z(s_j) \) can be represented as follows:

\[
(3.2) \quad Z(s_j) = \begin{cases} q(t(s_j)), & \text{if } I_z(s_j) = 1 \\ 0, & \text{if } I_z(s_j) = 0 \end{cases}.
\]

In a similar fashion to our emulation framework, we set up our model for the transformed thickness \( t(s_j) \) and the logit of \( I_z(s_j) \) denoted as \( \lambda(s_j) \). Following Chang et al. (2016a) we set up the following model to link it to the logit for the model output at the best setting \((\gamma(\theta^*, s_j))\) while accounting for data-model discrepancy:

\[
(3.3) \quad \lambda(s_j) = \gamma(\theta^*, s_j) + \alpha(s_j),
\]

where \( \theta^* \) is the input parameter setting that gives the ‘best’ match between model output and observational data, and \( \alpha(s_i) \) is a spatially correlated discrepancy term, normally distributed, representing the sum of spatially correlated parts of the model structural error (i.e. model mis-representation of the reality) and the observational error. The independent part of the errors are automatically taken care by the conditionally independent Bernoulli distribution for each location \( s_j \) given the logit \( \lambda(s_j) \) as in a usual generalized linear model setting.

The model for \( t(s_j) \) needs to be defined only for the locations with \( I_z(s_j) = 1 \). Let \( m = \sum_{j=1}^{p} I_z(s_j) \) be the number of spatial locations with a positive observed thickness. Without loss of generality, we assume that the observed thicknesses at the first \( m \) locations \( Z^+ = [Z(s_1), \ldots, Z(s_m)] \) are positive while the rest \( Z(s_{m+1}), \ldots, Z(s_p) \) are 0. For \( s_1, \ldots, s_m \), we use the following model for the transformed thickness:

\[
(3.4) \quad t(s_j) = \eta(\theta^*, s_j) + \delta(s_j) + \epsilon(s_j),
\]
where the random variables $\delta = [\delta(s_1), \ldots, \delta(s_m)]^T \sim N(0, \Sigma)$ and $\epsilon = [\epsilon(s_1), \ldots, \epsilon(s_m)]^T \sim N(0, \sigma^2 I_m)$ respectively represent the spatially correlated part and the i.i.d. part of the data-model discrepancy, which together represent the sum of the model structural errors and the observational errors. The discrepancy covariance $\Sigma$ reflects the spatial dependence among $\delta(s_1), \ldots, \delta(s_m)$.

It is worth mentioning that we do not follow the assumption in Kennedy and O’Hagan (2001) that the model structural errors and the observational errors can be separated into a dependent error component and an i.i.d. error component. This is because there is no reason to believe that the observational errors are i.i.d. and hence the two sources of errors are not distinguishable in most geoscientific applications. Rather we interpret the data-model discrepancy as ‘all effects that make the model output and observational data different’.

Moreover, we use a common link function $q(x)$ in (3.1) and (3.2) so that the latent processes for the emulator $t(s)$ and for the calibration model $\lambda(s)$ have the same interpretation; this allows the emulator for $h(\theta, s)$ to be directly used in modeling $\lambda(s)$.

The model in (3.3) assigns the following Bernoulli distribution for $I_z(s_j)$ (conditionally on the value of $\theta^*$ and discrepancy $\alpha(s_j)$):

$$P(I_z(s_j) = x|\gamma(\theta^*, s_j), \alpha(s_j)) = \left( \frac{\exp(\gamma(\theta^*, s_j) + \alpha(s_j))}{1 + \exp(\gamma(\theta^*, s_j) + \alpha(s_j))} \right)^x \left( 1 - \frac{\exp(\gamma(\theta^*, s_j) + \alpha(s_j))}{1 + \exp(\gamma(\theta^*, s_j) + \alpha(s_j))} \right)^{1-x}.$$ 

Given this distribution for $I_z(s_j)$, we can view the specification in (3.2) as a mixture model with density:

$$f(Z(s_j)|\eta(\theta^*, s_j), \delta(s_j), \sigma^2, \gamma(\theta^*, s_j), \alpha(s_j))$$

$$= \left| \frac{\partial t(s_j)}{\partial Z(s_j)} \right| f(t(s_j)|\eta(\theta^*, s_j), \delta(s_j), \sigma^2) P(I_z(s_j) = 1|\gamma(\theta^*, s_j), \alpha(s_j))$$

$$+ D_0(Z(s_j)) P(I_z(s_j) = 0|\gamma(\theta^*, s_j), \alpha(s_j))$$

for all locations $s_1, \ldots, s_p$, where the density function $f(t(s_j)|\eta(\theta^*, s_j), \delta(s_j), \sigma^2)$ is given by (3.4) and $D_0$ is the Dirac delta function. Since the density in (3.5) can be re-written as

$$f(t(s_j)|\eta(\theta^*, s_j), \delta(s_j), \sigma^2)$$

$$= \begin{cases} f(t(s_j)|\eta(\theta^*, s_j), \delta(s_j), \sigma^2) P(I_z(s_j) = 1|\gamma(\theta^*, s_j), \alpha(s_j)) & \text{if } I_z(s_j) = 1, \\ P(I_z(s_j) = 0|\gamma(\theta^*, s_j), \alpha(s_j)) & \text{if } I_z(s_j) = 0, \end{cases}$$

and $Z(s_1), \ldots, Z(s_p)$ are conditionally independent given the relevant parameters, the likelihood for $Z$ can be factorized as follows:

$$\mathcal{L}(Z|\eta^+(\theta^*), \delta, \sigma^2, \gamma(\theta^*), \alpha)$$

$$\propto \prod_{j=1}^m f(t(s_j)|\eta(\theta^*, s_j), \delta(s_j), \sigma^2) P(I_z(s_j) = 1|\gamma(\theta^*, s_j), \alpha(s_j))$$

$$\times \prod_{j=m+1}^p P(I_z(s_j) = 0|\gamma(\theta^*, s_j), \alpha(s_j)),$$

$$= \mathcal{L}_1(Z^+|\eta^+(\theta^*), \delta, \sigma^2) \mathcal{L}_2(I_z(s_1), \ldots, I_z(s_p)|\gamma(\theta^*), \alpha),$$
where
\[
L_1(Z^+|\eta^+(\theta^*), \delta, \sigma^2) = \prod_{j=1}^{m} f(t(s_j)|\eta(\theta^*, s_j), \delta(s_j), \sigma^2),
\]
\[
L_2(I_z(s_1), \ldots, I_z(s_p)|\gamma(\theta^*), \alpha) = \prod_{j=1}^{m} P(I_z(s_j) = 1|\gamma(\theta^*, s_j), \alpha(s_j)) \times \prod_{j=m+1}^{p} P(I_z(s_j) = 0|\gamma(\theta^*, s_j), \alpha(s_j)).
\]

Here \(\eta^+(\theta^*)\) is the vector of emulated process for all positive \(Z(s_j)\)'s (i.e. \(\eta^+(\theta^*) = [\eta(\theta^*, s_1), \ldots, \eta(\theta^*, s_m)]^T\)), and \(\alpha = [\alpha(s_1), \ldots, \alpha(s_p)]^T\). The Jacobian factors \(\frac{\partial \eta(s_j)}{\partial Z(s_j)}\) are omitted as they do not depend on any model parameters.

Interestingly the likelihood function that started from the mixture distribution-like specification in (3.5) leads to a factored likelihood in (3.6), which shows that the likelihood for \(Z\) can be factored into two parts, one for the positive observations \(Z^+\) and the other for the indicator variables at all locations \(I_z(s_1), \ldots, I_z(s_p)\). This has an important implication for inference on \(\theta^*\): utilizing the ice thickness pattern for calibration is essentially using the additional information from the positive ice thickness values \(Z^+\) on top of the binary spatial pattern of ice presence \((I_z(s_1), \ldots, I_z(s_p))\) in calibration. We will show how this added information improves our inference on the input parameter \(\theta^*\) in both the simulated and the real data examples in Section 5 below.

Note that this formulation does not necessarily require independence between \(Z^+\) and \(I_z(s_1), \ldots, I_z(s_p)\), because dependence can easily be specified through dependence between \(\eta(\theta^*, s_j)\) and \(\gamma(\theta^*, s_j)\) or \(\delta(s_j)\) and \(\alpha(s_j)\). This is how we impose dependence between \(Z^+\) and \(I_z(s_1), \ldots, I_z(s_p)\) in our formulation (see Section 4.2 below).

3.4. Computational and Inferential Challenges. The basic framework described in Section 3.3 may face some computational and inferential challenges when the model output and the observational data are in the form of high-dimensional spatial data (i.e. \(p\) is large) as in our PSU-3D Ice model calibration problem: First, inference based on the formulations described in Sections 3.2 and 3.3 requires to handle a large number of latent variables for the logits. To be more specific the number of latent variables in the emulation step is \(n \times p\) and this translates to \(499 \times 3,182 \approx 1.6\) million variables to infer for our problem. In the calibration step, while the number of latent variables is much smaller than that in the emulation step (\(2p = 6,364\)), the number of available data points (\(p\)) is much smaller than the number of latent variables (\(2p\)) and hence the problem is in fact ill-posed. Second, the size of data for height patterns from the model output is still large even when we consider only those at \(\theta_i\) and \(s_j\) with \(I_y(\theta_i, s_j) = 1\). In our calibration problem, the number of \((\theta_i, s_j)\) combinations with \(I_y(\theta_i, s_j) = 1\) is about 690,000 and this makes the standard Gaussian process emulation approach computationally infeasible because of the well-known computational issue with a large covariance matrix (see, e.g., Heaton et al., 2018).

4. Dimension Reduction-Based Approach. We mitigate the aforementioned challenges due to high-dimensional spatial data using the likelihood-based principal component
analysis (PCA) methods (Tipping and Bishop, 1999). Unlike the singular value decomposition-based PCA, the likelihood-based PCA can easily handle non-Gaussian data or partially observed data and hence is highly suitable for our problem.

Salter et al. (2019) recently has cautioned about possible issues regarding use of principal components in calibration—if the overall range for model output does not cover the range for observational data, calibration based on principal components can yield nonsensical results. Salter et al. (2019) has also proposed an optimal basis approach that can provide a solution in such situation. Chang et al. (2014) and Chang et al. (2016b) also discuss possible issues in a similar vein from the viewpoint of constructing discrepancy terms. Since the model runs and observational data discussed in Section 2 does not have such issues, we choose not to implement the optimal basis approach by Salter et al. (2019). (See Section 5.5 below for our detailed discussion on this point.)

4.1. Emulation Based on Likelihood-based Principal Component Analysis. We first introduce our dimension-reduced emulation method for binary spatial patterns, which is previously proposed by Chang et al. (2016a). Let \( \Gamma = [\gamma(\theta_1), \ldots, \gamma(\theta_n)]^T \) be a matrix of logits for the binary patterns \( \{I_y(\theta_i, s_j)\} \) (\( i = 1, \ldots, n \) and \( j = 1, \ldots, p \)) for the existing model runs. The rows of \( \Gamma \) correspond to the design points in input parameter settings \( \theta_1, \ldots, \theta_n \) while the columns are for different spatial locations \( s_1, \ldots, s_p \). We apply logistic principal component analysis (LPCA) (Lee et al., 2010) to decompose the logit matrix \( \Gamma \) in the following way:

\[
\Gamma = 1_n \mu^T + W K_w^T,
\]

where \( \mu \) is the \( p \times 1 \) mean vector for the spatial locations \( s_1, \ldots, s_p \) (i.e. the column means of \( \Gamma \)), \( W \) is the \( n \times J_w \) logistic principal component (LPC) score matrix, and \( K_w \) is the \( p \times J_w \) LPC matrix with a pre-specified number of principal components \( J_w \geq 1 \). The rows of \( W = [w(\theta_1), \ldots, w(\theta_n)]^T \) correspond to the logits for different input parameter settings where \( w(\theta) = [w_1(\theta), \ldots, w_{J_w}(\theta)]^T \) denotes a vector of the LPC scores at \( \theta \). The parameters in (4.1) (\( \mu, W, \) and \( K_w ) \) can be estimated by maximizing the corresponding likelihood function for these parameters given the binary patterns \( \{I_y(\theta_i, s_j)\} \) for existing model runs using the minorization and maximization (MM) algorithm. We predict the logits \( \gamma(\theta) \) at any untried setting \( \theta \) by predicting the corresponding LPC scores \( w(\theta) \). Each score \( w_k(\theta) \) (for \( k = 1, \ldots, J_w \)) can be predicted separately using a GP emulator. (See Section S3 in SI for details.) We denote the resulting emulated process of LPC scores at \( \theta \) as \( \psi(\theta) = [\psi_1(\theta), \ldots, \psi_{J_w}(\theta)]^T \).

We also apply a likelihood-based PCA method for data with missing values to build an emulator for the ice-thickness patterns. For \( \theta_i \) and \( s_j \) with \( I_y(\theta_i, s_j) = 1 \) we assume the following model for dimension reduction:

\[
h(\theta_i, s_j) = \sum_{l=1}^{J_u} k_{u, jl} u_l(\theta_i) + e_{ij}
\]

with \( e_{ij} \sim \text{i.i.d } N(0, \sigma_e^2) \) (\( \sigma_e^2 > 0 \)), the principal component (PC) loading \( k_{u, jl} \) (\( j = 1, \ldots, p \) and \( l = 1, \ldots, J_u \)) and the PC score \( u_l(\theta_i) \) (\( i = 1, \ldots, n \) and \( l = 1, \ldots, J_u \)). Again \( J_u \geq 1 \) is the pre-determined number of principal components being used for our dimension reduction.
ICE MODEL CALIBRATION USING SEMI-CONTINUOUS SPATIAL DATA

344 This is essentially PCA with missing values and therefore the PC loadings and scores can be 345 estimated via EM algorithm (Stacklies et al., 2007). We denote the resulting \( p \times J \) loading 346 matrix by \( K_u \), with \((i,j)\)th element given by \( k_{u,ij} \). In a similar manner to the problem of 347 emulating logits we predict the latent variables for the thickness \( h(\theta, s_j) \) at any untried 348 setting \( \theta \) and location \( s_j \) with a positive thickness value by predicting the corresponding 349 principal component scores \( u_l(\theta) = [u_{l1}(\theta), \ldots, u_{J_n}(\theta)]^T \).

350 Again we build an emulator for each principal component separately using a GP emulator 351 with the following exponential covariance function:

\[
(4.3) \quad \text{Cov}(u_l(\theta), u_l(\theta')) = \zeta_{u,l} I(\theta = \theta') + \kappa_{u,l} \exp\left(-\sum_{b=1}^{d} \frac{|\theta_b - \theta'_b|}{\phi_{u,l,b}}\right)
\]

352 for any two input parameter settings \( \theta \) and \( \theta' \) where \( \zeta_{u,l} > 0 \) is the nugget, \( \kappa_{u,l} > 0 \) is the 353 partial sill, and \( \phi_{u,l1}, \ldots, \phi_{u,ld} > 0 \) are the range parameters. To incorporate information 354 from the binary pattern we use the following mean function for the \( l \)th principal component:

\[
(4.4) \quad E(u_l(\theta_i)|w_1(\theta_i), \ldots, w_{J_n}(\theta_i)) = \sum_{k=1}^{J_n} g_{lk}(w_k(\theta_i)),
\]

355 where the function \( g_{lk} \) is given by a natural spline regression model whose degrees of freedom 356 is determined through cross-validation (Hastie, 1992). We let \( \beta_{lk} \) be the vector of coefficients 357 for \( g_{lk}(\cdot) \), whose dimensionality is the same as the degrees of freedom of \( g_{lk} \). To construct the 358 GP emulator we find the posterior modes of the covariance parameters (denoted as \( \hat{\zeta}_{u,l}, \hat{\kappa}_{u,l} \) 359 and \( \hat{\phi}_{u,l1}, \ldots, \hat{\phi}_{u,ld} \)) and the parameters for the spline functions (denoted as \( \hat{\beta}_{l1}, \ldots, \hat{\beta}_{lJ_n} \) 360 for each \( l \)th principal component separately using the robust Bayesian procedure (Gu et al., 361 2018, 2019) as we do for the LPC scores above. When we predict \( u_l(\theta) \) for any untried 362 setting \( \theta \not\in \{\theta_1, \ldots, \theta_n\} \), we replace \( w_k(\theta) \) with \( E(w_k(\theta)|w_1(\theta_1), \ldots, w_{J_n}(\theta_n)) \) given by the 363 Gaussian process emulator described above since \( w_k(\theta) \) is not available if \( \theta \not\in \{\theta_1, \ldots, \theta_n\} \). We let \( \xi(\theta) = [\xi_1(\theta), \ldots, \xi_{J_n}(\theta)]^T \) denote the resulting emulated process for \( u(\theta) \).

For any untried input parameter setting \( \theta \), we can predict the ice thickness pattern from 366 our computer model in the following two steps: (i) We first predict the logits of ice-no ice 367 patterns \( \gamma(\theta) = K_u \psi(\theta) \), and (ii) for each location \( s_j \) with \( \gamma(\theta, s_j) > 0 \) the predicted 368 thickness is given as \( q \left( \sum_{l=1}^{J_n} k_{u,lj} u_l(\theta) \right) \). Note, however, that the thresholding of the logits 369 at 0 is needed only for evaluating emulation performance (such as generating predicted 370 patterns for visual evaluation) and is not used in our actual calibration procedure.

In the calibration step discussed below, we fix the emulator parameters at their posterior 371 modes except for the partial sill parameters for \( \xi, \kappa_u = [\kappa_{u,1}, \ldots, \kappa_{u,J_n}] \). The partial sill 372 parameters for \( \xi \) will be re-estimated along other parameters in the calibration model to 373 account for any possible discrepancies in scale (see e.g., Bhat et al., 2012; Chang et al., 374 2014, 2015, 2016b, for similar approaches ). However, the partial sills for \( \psi \) will be fixed at 375 their MLEs without being re-estimated in the calibration stage because the binary patterns 376 usually do not have enough information for the scale parameters of the latent variables 377 and hence re-estimation for the partial sill parameters often cause identifiability issues as 378 discussed in Chang et al. (2016a).
4.2. Calibration Using Basis Representation. Using the emulators for principal components ($\psi$ and $\xi$) described in the previous section we modify the basic calibration framework introduced in Section 3.3 to set up a computationally efficient calibration method. We now rewrite the model for $t(s_j)$ in (3.4) as

$$
t(s_j) = \sum_{l=1}^{J_u} k_{u,jl}\xi_l(\theta^*) + \sum_{k=1}^{J_r} k_{r,jk}r_k + \epsilon_j
$$

for $j = 1, \ldots, m$, where $k_{r,jk}$ is the $(j,k)$th element of an $m \times J_r$ discrepancy basis matrix $K_r$, $r_1, \ldots, r_{J_r}$ $\sim$ i.i.d. $N(0, \sigma^2_r)$ are the random coefficients with $\sigma^2_r > 0$ for $K_r$, and $\epsilon_j \sim N(0, \sigma^2_\epsilon)$ is the i.i.d. observational error with $\sigma^2_\epsilon > 0$. The terms $\sum_{l=1}^{J_u} k_{u,jl}\xi_l(\theta^*)$ and $\sum_{k=1}^{J_r} k_{r,jk}r_k$ are respectively the basis representations of $\eta(\theta^*, s_j)$ and $\delta(s_j)$ in (3.4) given by our formulation.

We also rewrite the model for the logits $\lambda$ for $Z$ in (3.3) using a similar basis representation as follows:

$$
\lambda = \mu + K_u\psi(\theta^*) + K_v\nu,
$$

with the logistic principal component basis matrix $K_u$, and a $p \times J_v$ discrepancy basis matrix $K_v$ and its corresponding coefficients $\nu = [v_1, \ldots, v_{J_v}]^T \sim N(0, \sigma^2_\nu I_{J_v})$ with $\sigma^2_\nu > 0$. This calibration model formulation has been previously proposed in Chang et al. (2016a). We model the dependence between the coefficients of the discrepancy terms $\nu = [v_1, \ldots, v_{J_v}]^T$ and $r = [r_1, \ldots, r_{J_r}]^T$ through a $J_v \times J_r$ cross correlation matrix $R$, whose $(i,j)$th element $\rho_{ij}$ is the correlation between $v_i$ and $r_j$.

The discrepancy basis matrices $K_r$ and $K_v$ need to be carefully specified to avoid possible identifiability issues between the effects of input parameters and the discrepancy. For our analysis described in Section 5, the discrepancy basis $K_v$ and $K_r$ are both defined using exponential kernels with knot locations marked as red dots in Figure S5 and their range is defined as 80km. They are defined based on our prior knowledge that (i) the used PSU-3D ice model runs with 20km resolution has more model errors in the areas with small-scale observed features and ice-flow gradients, and (ii) there could be more observational errors as well in the marked regions. In the left area covered with knots is the outlet of Pine Island Glacierots, there is a narrow channel between Thurston island and the mainland. Moreover, as Figures 11 and 12 in Fretwell et al. (2013) suggest, the thickness observations around the Thurston Island area have large observational errors, due to the large uncertainty in the bed rock elevation. Located in the right area with knots is the outlet of Pine Island Glacier, a narrow but fast moving glacier stream. We expect that magnitudes of the combined observational and model structural errors for the ice thickness are around 500-1000m, and this prior knowledge is reflected in our simulated example in Section 5.4.

The formulation in (4.5) implies that the role of thickness emulator $\sum_{l=1}^{J_u} k_{u,jl}\xi_l(\theta_i)$ is restricted to capturing the ice thickness when the observed thickness is greater than 0. The information from zero thickness patterns are separately handled by the process $\lambda(s)$. This makes consideration of ‘missing mechanism’ in PCA with missing values in (4.2) unimportant because the ‘missed thickness values’ at zero locations do not play any role in inference based on the calibration models in (4.5) and (4.6).
Note that our formulation does not strictly require that the model output and the observational data have the same dimensions. In some applications some pixels contained in the model output \( Y(\theta) \) might be missing in the observational data \( Z \) because obtaining observational data can be harder at certain locations. In such a case one can simply re-adjust the basis matrices for the emulators \( (K_u \text{ and } K_w) \) accordingly by deleting the rows corresponding to the missing locations.

4.3. Bayesian Inference. Given the above formulation we conduct Bayesian inference on \( \theta^* \) and other parameters in the model. While using non-Bayesian inference might be possible as well, we choose to use a Bayesian method as it provides a quite straightforward way to quantify the uncertainty in \( \theta^* \) while account for other sources of uncertainties despite the complexity of our model specification. All the emulator parameters for \( \psi(\theta^*) \) and \( \xi(\theta^*) \) except for the sill parameters for \( \psi(\theta^*) \) \( (\kappa_u) \) are fixed at their MLE computed in the emulation stage. This two stage approach with fixed emulator parameters is helpful to reduce identifiability issues in calibration (Bayarri et al., 2007; Bhat et al., 2012; Chang et al., 2014).

Likelihood. In a similar fashion to the specification in (3.5) the representations in (4.5) and (4.6) lead to a density function based on a mixture model. The likelihood function for the mixture model conditional on the emulated process \( \xi \) and \( \psi \) now becomes

\[
f(Z(s_j)|\xi(\theta^*), r, \sigma_v^2, \psi(\theta^*), v) = \left| \frac{\partial t(s_j)}{\partial Z(s_j)} \right| f(Z(s_j)|\xi(\theta^*), r, \sigma_v^2) P(I_z(s_j) = 1|\psi(\theta^*), v) + D_0(Z(s_j)) P(I_z(s_j) = 0|\psi(\theta^*), v)
\]

for all locations \( s_1, \ldots, s_p \), where the density function \( f(\xi(\theta^*), r, \sigma_v^2) \) is for the case with \( I_z(s_j) = 1 \) in (4.5). As a result we have the following likelihood function for \( Z \):

\[
\mathcal{L}(Z|\xi(\theta^*), r, \sigma_v^2, \psi(\theta^*), v) \propto \prod_{j=1}^{m} f(Z(s_j)|\xi(\theta^*), r, \sigma_v^2) P(I_z(s_j) = 1|\psi(\theta^*), v) \\
\times \prod_{j=m+1}^{p} P(I_z(s_j) = 0|\psi(\theta^*), v).
\]

where

\[
\mathcal{L}_1(Z^+|\xi(\theta^*), r, \sigma_v^2) = \prod_{j=1}^{m} f(Z(s_j)|\xi(\theta^*), r, \sigma_v^2), \\
\mathcal{L}_2(I_z(s_1), \ldots, I_z(s_p)|\psi(\theta^*), v) = \prod_{j=1}^{m} P(I_z(s_j) = 1|\psi(\theta^*), v) \prod_{j=m+1}^{p} P(I_z(s_j) = 0|\psi(\theta^*), v).
\]

We have a similar factorization as in (3.6) with one factor for the positive observations \( Z^+ \) and the other for the binary variables at all locations \( I_z(s_1), \ldots, I_z(s_p) \).
Prior. To complete the Bayesian model specification, we assign the following priors for the model parameters \((\theta^*, \nu, \sigma_r^2, \sigma_v^2, \kappa_u, R)\) in our calibration step:

\[
\begin{align*}
  v_j | \sigma_v^2 & \sim N(0, \sigma_v^2) \quad j = 1, \ldots, J_v; \\
  \sigma_r^2 & \sim IG(\alpha_r, \beta_r); \\
  \kappa_{u,j} & \sim IG(5, 6\kappa_{u,j}) \quad j = 1, \ldots, J_u; \\
  \theta^* & \sim d\pi(\theta^*)
\end{align*}
\]

where \(f(R)\) is a uniform distribution within the range that \(I_J - RR^T\) is positive definite, i.e., \(f(R) \propto I(I_J - RR^T)\) is positive definite. The shape \((\alpha_r)\) and the scale \((\beta_r)\) parameters for the variance of thickness discrepancy \(\sigma_r^2\) is set to be \(\alpha_r = 50\) and \(\beta_r = 500\alpha_r/\max(\text{diag}(\sigma_r^2K_rK_r^T))\) where \(\text{diag}(\sigma_r^2K_rK_r^T)\) means the diagonal elements of \(\sigma_r^2K_rK_r^T\). This encourages the variance of the resulting discrepancy term \(K_rr\) (i.e. the diagonal elements of the covariance matrix \(\sigma_r^2K_rK_r^T\)) to be around 500² or slightly less, to reflect our knowledge on observational errors on the thickness measurements Fretwell et al. (2013) and the model structural errors in the area covered by the discrepancy term (expected to be around 500-1000m). The mode of the resulting inverse-Gamma distribution is around 300². Similarly, the prior for \(\sigma_v^2\) is designed to have a mode around 10² with \(\alpha_v = 50\) and \(\beta_v = 100²(\alpha_v + 1)\), to reflect our belief that the magnitudes of differences between the model output and the observational data are 50-200m in the region other than the high-discrepancy area shown in Figure S5, covered by the discrepancy term \(K_rr\).

As per the prior for \(\theta^*\), we use prior densities inspired by previous calibration results for OCFAC, CALV, CRH, and TAU based on modern and paleo-data (‘Simple Method’ in Figure 2 by Pollard et al., 2016). The informative priors used here are

\[
\begin{align*}
  OCFAC & \sim TN(0.75, 0.1^2, 0, 1); \\
  CALV & \sim TN(0.3, 0.15^2, 0, 1); \\
  CRH & \sim TN(0.8, 0.1^2, 0, 1); \\
  TAU & \sim TN(0.4, 0.15^2, 0, 1),
\end{align*}
\]

where \(TN(\mu, \sigma^2, \text{low}, \text{upper})\) denotes a truncated normal distribution with the mean parameter, \(\mu\), variance, \(\sigma^2\), and the lower and upper bounds, \(\text{lower}\) and \(\text{upper}\). Note that these priors are specified with respect to the transformed \([0,1]\) scale, not to the original scale. The mean and the standard deviation for OCFAC and CRH are chosen to reflect our prior belief that the values of these parameters are distributed in the higher part of the parameter space and we are a little more uncertain about these two parameters than CALV and TAU. The prior distributions for CALV and TAU are chosen to reflect our belief that their values are distributed in the lower or middle part of the parameter space. For the other six parameters, we assign independent uniform priors within the ranges specified in Table S1, which are determined by the domain knowledge in the previous literature (Pollard and DeConto, 2012). Please note that these prior densities for the input parameters are used for both the calibration based on the full ice thickness patterns and the calibration based only on the binary patterns explained below.

We have specified a vague prior on \(\sigma_v^2\) with a shape parameter of 2 to avoid possible issues with improper posterior and computational instability (Berger et al., 2001). We have noticed that the posterior analysis is largely insensitive to the choice of the prior hyper-parameters (see Section 5.5 for further discussion on this). For the re-estimated partial
sill parameters \( \kappa_{u,1}, \ldots, \kappa_{u,J_u} \), we assigned a slightly informative prior to encourage them to have values around their MLEs estimated in the emulation stage. For the re-estimated partial sill parameters \( \kappa_{u,1}, \ldots, \kappa_{u,J_u} \), we assigned a slightly informative prior to encourage them to have values around their MLEs estimated in the emulation stage.

**Posterior.** The above specification of likelihood and prior lead to a posterior whose density can be factorized as follows:

\[
\pi(\theta^*, \mathbf{v}, \sigma^2_r, \sigma^2_{\epsilon}, \sigma^2_u, \kappa_u, \mathbf{R} | \mathbf{Z}) \propto \pi_1(\sigma^2_r, \sigma^2_{\epsilon}, \kappa_u, \mathbf{R} | \theta^*, \mathbf{v}, \mathbf{Z}^+) \\
\times \pi_2(\theta^*, \mathbf{v}, \sigma^2_{\epsilon} | I_\epsilon(s_1), \ldots, I_\epsilon(s_p)).
\]  

The first part on the right-hand side is based on the likelihood for \( \mathbf{Z}^+ (L_1) \) and the relevant priors and obtained by

\[
\pi_1(\sigma^2_r, \sigma^2_{\epsilon}, \kappa_u, \mathbf{R} | \theta^*, \mathbf{v}, \mathbf{Z}^+) \propto \int \int L_1(\mathbf{Z}^+ | \mathbf{\xi}(\theta^*), \mathbf{r}, \sigma^2_r) f(\mathbf{\xi}(\theta^*) | \theta^*, \mathbf{K}_u) f(\mathbf{r} | \sigma^2_r, \mathbf{v}) d\mathbf{r} d\mathbf{\xi}
\]

\[
\times f(\sigma^2_r) f(\kappa_u) f(\sigma^2_{\epsilon}) f(\mathbf{R})
\]

\[
= \mathcal{L}_1^*(\mathbf{Z}^+ | \theta^*, \sigma^2_r, \sigma^2_{\epsilon}, \kappa_u, \mathbf{v}, \mathbf{R}) f(\sigma^2_r) f(\kappa_u) f(\sigma^2_{\epsilon}) f(\mathbf{R}),
\]

where \( f(\sigma^2_r) \), \( f(\theta^*) \), \( f(\kappa_u) \), \( f(\sigma^2_{\epsilon}) \), and \( f(\mathbf{R}) \) are the prior densities (defined above) and the marginal likelihood \( \mathcal{L}_1^* \) can be written as

\[
\mathcal{L}_1^*(\mathbf{Z}^+ | \theta^*, \sigma^2_r, \sigma^2_{\epsilon}, \kappa_u, \mathbf{v}, \mathbf{R}) \propto |\Sigma_+|^{-1/2} \exp \left[ -\frac{1}{2} (\mathbf{q}^{-1}(\mathbf{Z}^+) - \mathbf{\mu}_+)^T \Sigma_+^{-1}(\theta^*) (\mathbf{q}^{-1}(\mathbf{Z}^+) - \mathbf{\mu}_+) \right],
\]

with \( \mathbf{q}^{-1}(\mathbf{Z}^+) = [\mathbf{q}^{-1}(\mathbf{Z}(s_1)), \ldots, \mathbf{q}^{-1}(\mathbf{Z}(s_m))]^T \). The mean and covariance of \( \mathbf{q}^{-1}(\mathbf{Z}^+) \) are given by

\[
\mathbf{\mu}_+ = \mathbf{K}_{+,u} \mathbf{\mu}_\xi(\theta^*) - \mathbf{K}_r \mathbf{\mu}_r|\mathbf{v}
\]

\[
\Sigma_+ = [\mathbf{K}_{+,u} \mathbf{K}_r] \Sigma_{\xi,r} [\mathbf{K}_{+,u} \mathbf{K}_r]^T + \sigma^2_{\epsilon} \mathbf{I}_m.
\]

Here \( \mathbf{\mu}_\xi(\theta^*) \) is the mean of the emulated process \( \mathbf{\xi}(\theta^*) \) and \( \mathbf{K}_{+,u} \) is an \( m \times J_y \) matrix created by collecting the first \( m \) rows of \( \mathbf{K}_u \); \( \Sigma_{\xi,r} \) is a block diagonal matrix defined as

\[
\Sigma_{\xi,r} = \begin{pmatrix}
\Sigma_{\xi} & 0 \\
0 & \Sigma_{r|\mathbf{v}}
\end{pmatrix},
\]

where \( \Sigma_{\xi} \) is a \( J_u \times J_u \) diagonal matrix whose diagonal elements are the conditional variances of \( \xi_1(\theta^*), \ldots, \xi_{J_u}(\theta^*) \) from the GP emulators defining \( \mathbf{\xi}(\theta^*) \); \( \mathbf{\mu}_r|\mathbf{v} \) and \( \Sigma_{r|\mathbf{v}} \) are the conditional mean and variance of \( \mathbf{r} \) given \( \mathbf{v} \) defined as

\[
\mathbf{\mu}_r|\mathbf{v} = \frac{\sigma_r}{\sigma_u} \mathbf{R} \mathbf{v},
\]

\[
\Sigma_{r|\mathbf{v}} = \sigma^2_r (\mathbf{I}_r - \mathbf{R} \mathbf{R}^T) .
\]
The computational cost for finding the inverse and the determinant of this covariance matrix can be significantly reduced using the Sherman-Woodbury-Morrison formula (Woodbury, 1950) and the determinant formula (Harville, 2008). See Appendix A for further details.

The second part of the posterior density is given as

$$\pi_2(\theta^*, v, \sigma_v^2|I_z(s_1), \ldots, I_z(s_p)) \propto \mathcal{L}_2(I_z(s_1), \ldots, I_z(s_p)|\psi(\theta^*), v) f(\psi(\theta^*)|\theta^*) f(\theta^*) f(v|\sigma_v^2) f(\sigma_v^2),$$

with the prior densities $f(\psi(\theta^*)|\theta^*)$, $f(\theta^*)$, $f(v|\sigma_v^2)$, and $f(\sigma_v^2)$. The formulation for the logits in (4.6) leads to the following likelihood function for $I_z(s_1), \ldots, I_z(s_p)$:

$$\mathcal{L}_2(I_z(s_1), \ldots, I_z(s_p)|\psi(\theta^*), v) \propto \prod_{j=1}^{p} \left[ \left( \frac{\exp(\lambda(s_j))}{1 + \exp(\lambda(s_j))} \right) I_z(s_j) \left( \frac{1}{1 + \exp(\lambda(s_j))} \right)^{1-I_z(s_j)} \right],$$

where the logits $\lambda(s_1), \ldots, \lambda(s_p)$ are determined by $\psi(\theta^*)$ and $v$ through the basis representation in (4.6). The prior for $\psi(\theta^*)$ is given by the Gaussian process emulator with the mean and covariance respectively given in (4.3) and (4.4) and hence has the following multivariate normal density:

$$f(\psi(\theta^*)|\theta^*) \propto |\Sigma_{\psi}(\theta^*)|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} [\psi(\theta^*) - \mu_{\psi}(\theta^*)]^T \Sigma_{\psi}^{-1}(\theta^*) [\psi(\theta^*) - \mu_{\psi}(\theta^*)] \right),$$

where $\mu_{\psi}(\theta^*)$ is a vector of conditional means given by the Gaussian process emulators for $\mu(\theta^*); \Sigma_{\psi}(\theta^*)$ is a diagonal matrix whose diagonal elements are given by the conditional variance from the same Gaussian process emulators.

The target input parameters and the other parameters can be inferred based on the posterior density in (4.7). To facilitate the Bayesian inference we can resort to MCMC methods, which, in our case, require sampling from the posterior distribution by using Metropolis within Gibbs sampling (Gilks et al., 1995; Gelfand and Smith, 1990).

5. Application. We now discuss the results of applying our method to the problem of PSU-3D ice model calibration based on Bedmap2 data described in Section 2. As the first step, we have built a dimension-reduced emulator described in Section 4, which takes about 3 hours on a single high-performance core if implemented in an R code. While further speed-up is possible by switching to a faster programming language or utilizing parallel computing we have decided not to pursue such an effort as the current implementation is fast enough for our purpose.

5.1. Selection of the Number of PCs. To choose the number of principal components for the likelihood-based PCA, we use reduction in deviance as the metric. The principal components used here are likelihood-based ones, and hence we cannot rely on the eigenvalues as in the standard principal component analysis with singular value decomposition. For the logistic principal components, the explained deviance is defined as

$$(5.1) \quad \text{explained deviance} = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{p} [I_y(\theta_i, s_j) \log(\hat{p}_{ij}) + (1 - I_y(\theta_i, s_j)) \log(1 - \hat{p}_{ij})]}{\sum_{i=1}^{n} \sum_{j=1}^{p} [I_y(\theta_i, s_j) \log(\hat{p}_{0j}) + (1 - I_y(\theta_i, s_j)) \log(1 - \hat{p}_{0j})]}. $$
where $\hat{p}_{ij} = \frac{\exp(\hat{\gamma}_{ij})}{1 + \exp(\hat{\gamma}_{ij})}$ is an estimator for $P(I(\theta_i, s_j) = 1)$ based on the fitted model with LPCA; $\hat{\gamma}_{ij}$ is an estimator for $\gamma_{ij}$, which is the $(i,j)$th element of $\Gamma$ in (4.1); $\hat{p}_{ij} = \sum_{i=1}^{n} I(\theta_i, s_j)/n$ is the estimator for $P(I(\theta_i, s_j) = 1)$ based on the ‘null model’, which assumes a constant probability $P(I(\theta_i, s_j) = 1) = p_{ij}$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, p$. For the principal components for thickness, the explained deviance is equivalent to the following explained variance defined as

\begin{equation}
\text{explained variance} = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{p} [Y(\theta_i, s_j) - \bar{Y}_{ij}]^2}{\sum_{i=1}^{n} \sum_{j=1}^{p} [Y(\theta_i, s_j) - \bar{Y}_{j}]^2}
\end{equation}

where $\bar{Y}_{ij}$ is the fitted value for $Y(\theta_i, s_j)$ from the model in (4.2), and $\bar{Y}_{j}$ is the estimator based on the null model which assumes a constant mean for all $\theta_1, \ldots, \theta_n$ at any $s_j$ ($j = 1, \ldots, p$).

We have calculated the explained deviance and variance for $J_w = 1, \ldots, 25$ and $J_u = 1, \ldots, 25$ (Figure S2). In addition to the metrics computed for the entire area, we have also computed the metrics for the area marked with a red box in Figure S3, in which the input parameters are expected to have complicated effects due to the complex terrain and interaction with ocean (‘focus area’ henceforth). For the logistic principal components, we set $J_w = 13$ to have more than 95% of explained deviance for the whole domain and 90% of reduction in deviation for the focus area. We use a lower threshold for the explained deviance in the focus area as achieving 95% would require too many logistic principal components (24 LPCs), which leads to poor prediction performance by the GP emulators due to the noisy behaviors of some of the non-leading components. For the principal components for the thickness, we set $J_w = 19$ to have more than 95% of explained variation for both the whole domain and for the focus area.

Here we are using far more principal components ($J_w = 13$ and $J_u = 19$) than typically recommended (around 5) according to the existing literature on principal component emulators (e.g. Higdon et al., 2008). To check if each principal component score is well predicted by the GP emulators, we have conducted 10-fold cross validation on each principal component and the results are summarized in Tables S2 and S3. The tables show the explained deviance or the explained variance for each LPC or PC and the correlation coefficients between the original and predicted values for each score. Higher the correlation coefficient, better the PC score is predicted. The results show that the predicted and the original scores have overall reasonably high correlation even for the non-leading principal components. This suggests that, for this application, even the non-leading PCs and LPCs selected in our analysis contain some signals that can be predicted based on input parameter values.

5.2. **Diagnostics of Emulation Performance.** To verify the performance of our emulator we conduct 10-fold cross-validation for the emulator. To be more specific, we have randomly left out 10% of the model runs and used the constructed emulator to predict the left out model outputs. This procedure is repeated 10 times (while choosing the model runs that are not selected previously) to compute the prediction accuracy for all model runs. Some example cases are shown in Figure 2. The cross-validation results show that our emulator can predict the left-out model outputs with a high accuracy, both in terms of the ice-no
ice binary patterns and the thickness patterns. The overall mean absolute error (MAE) for ice thickness prediction at the locations with positive thickness is about 231.05m (while the overall mean ice thickness at those locations is 2117m). The sensitivity (the percentage of left out runs where ice presence was correctly predicted) is 96.6% and the specificity (the percentage of left out runs where ice absence was correctly predicted) is 90.9%. When computing these scores the pixels with the same ice or no ice outcome for all the ensemble members are excluded from calculation.

We also examine how the prediction accuracy of the LPC and PC emulators changes as we vary the used number of model runs with \( n = 200, 300, 400, \) and 499, to see if Loeppky’s rule of thumb Loeppky et al. (2009) applies to our emulation model. The 10-fold cross-validation results summarized in Figure S4 show that, first, the prediction accuracy of the PC emulator notably improves as we add more model runs, indicating that Loeppky’s rule may not apply under the use of latent variables and the presence of missingness here. Second, the overall accuracy of the LPC emulator slightly increases as we add more runs, but the true positive and true negative rates behave differently. The true positive rate stays almost the same while the true negative rate increases by about 10% indicating that adding more runs does improve correct classification for the no ice area. This result and the comparison with thickness-only emulators above show that correctly predicting the zero areas in our problem requires more model runs and careful model choice than usual emulation efforts.

In addition, we compare the prediction performance of our approach with two alternatives: the original PCA-based emulator (Higdon et al., 2008; Chang et al., 2014) and the partial parallel emulator (Gu et al., 2016). Both are directly applied to the full thickness pattern without adjusting to account for zero output values, hoping that the emulator can capture the zero thickness by predicting them as near-zero values. To make these two emulators as flexible as possible, we use exponential covariance functions to capture the trends in the parameter space. See Section S6 for further details on these emulators. The comparison results are summarized in Table 1 and example comparison results are shown in Figure S6. The overall MAEs listed in Table 1 show that our method outperforms the other two alternatives. This is because, as shown in Figure S6, the two alternative methods often fail to predict a large zero thickness large area. If some changes in the parameter values lead to reduction in ice thickness below certain threshold, an abrupt disintegration of ice sheet in a certain part of study area can occur and this behavior cannot be captured by a stationary covariance structure even with the exponential covariance function.

For the ‘PCA only for Thickness’ emulator in Table 1, the number of PCs is selected based on the explained variance defined in 5.2 for both the overall region and the focal area in Figure S3, as described in Section S6. We also include the results based on 19 PCs for the ‘PCA only for Thickness’ emulator, to show that the difference between this emulator and the proposed method is not merely due to the number of PCs used here. The prediction accuracy is notably worse with both choices (10 PCs and 19 PCs).

Using the constructed emulators and the observational data we infer the best input parameter setting \( \theta^* \). We first verify our method using a synthetic data example in Section 5.4 and proceed to calibration using the real observations from Bedmap2 data in Section 5.5. In both cases we compare our current method (full approach henceforth) to the calibration results obtained using only the ice/no ice binary patterns (binary-only approach henceforth, originally presented in Chang et al., 2016a) to show the added value by fully utilizing the
ice thickness patterns in calibration.

5.3. Choice of Transformation Function and Emulation Performance. The success of this latent variable-based approach partially depends on the choice of the transformation function \( q \) to guarantee non-negativity without introducing a serious artifact due to transformation. While in the literature an exponential transformation is commonly used to enforce non-negativity, we found that the use of an exponential transformation imposes too much distortion in distribution and results in a poor emulation performance in our problem (MAE of about 453.795 m, about twice higher than that of our result). Therefore in this study we use the following link function that can ensure non-negativity with only a minimal distortion of data distribution:

\[
q(x) = \begin{cases} 
  x & \text{if } x > 1, \\
  \exp(x - 1) & \text{if } x \leq 1.
\end{cases}
\]

This function preserves the original pattern of ice thickness by setting \( h(\theta_i, s_j) = Y(\theta_i, s_j) \) for \( Y(\theta_i, s_j) > 1 \) m, while allowing the transformed variable to have negative values by setting \( h(\theta_i, s_j) = \log(Y(\theta_i, s_j)) + 1 \) for \( 0 < Y(\theta_i, s_j) \leq 1 \) m. This function also ensures a smooth transition at \( x = 1 \) because \( \frac{\partial q(x)}{\partial x} \) exists and has a value of one when \( x = 1 \).

One drawback of the above transformation is that the calibration of the ice thickness \( q(\eta(\theta^*, s) + \delta(s) + \epsilon) \) is different for ice thickness smaller than one meter and for ice thickness greater than one meter. More precisely, the calibration formulation is multiplicative for ice thickness of magnitude less than one meter and additive for ice thickness of greater or equal to one meter. However, for the observational data used in our application, the percentage of pixels with ice thickness lower than one meter is 0.01% for both simulated and observed data sets. This implies that our calibration process is in practice an additive calibration model. The modeled and observed thicknesses are compared directly without any transformation in most cases.

5.4. Calibration Using Synthetic Data. We now verify the performance of our calibration method using a synthetic data example. To generate a synthetic data set we choose the true input parameter setting and its corresponding output for ice thickness pattern as the assumed truth. We then superimpose generated errors to represent a possible data-model discrepancy in reality. We chose a model output whose input parameter values are not at the center of the cloud of design points to make the test more challenging. Based on the selected assumed truth, the synthetic data is generated as follows:

### Table 1

<table>
<thead>
<tr>
<th>Method</th>
<th># of LPCs</th>
<th># of PCs</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial Parallel</td>
<td>-</td>
<td>-</td>
<td>410.063</td>
</tr>
<tr>
<td>PCA only for Thickness</td>
<td>-</td>
<td>10</td>
<td>358.512</td>
</tr>
<tr>
<td>PCA only for Thickness</td>
<td>-</td>
<td>19</td>
<td>342.118</td>
</tr>
<tr>
<td>Proposed Emulator</td>
<td>13</td>
<td>19</td>
<td>231.050</td>
</tr>
</tbody>
</table>
Fig 2: Examples of 10-fold cross-validation results, showing selected original spatial patterns from PSU-3D ice model (left column) and the corresponding emulated patterns (right column). The comparison shows that our emulator can predict the original model output with high accuracy.
(a) We first generate the assumed logit $\lambda$ based on Equation (4.6), by computing the predicted mean of $\xi(\theta^*)$ at the assumed truth $\theta^*$ using the emulator constructed without the model run at $\theta^*$. We then sample the coefficients for the discrepancy term $K_v$, $v$ from $N(0,0.4^2I)$. The variance $0.4^2$ is chosen by trying on different values and selecting the one that yields a reasonably challenging discrepancy pattern. Then for each location $s_j$ we determine the ice presence and absence by sampling $I(s_j)$ from a Bernoulli random variable with a success probability of $\frac{\lambda_j}{1+\lambda_j}$, where $\lambda_j$ is the $j$th element of $\lambda$.

(b) Secondly, we superimpose $K_r r + \epsilon$ on $Y(\theta^*)$ at the assumed truth for $\theta^*$, where $r \sim N(0,300^2I)$ and $\epsilon \sim N(0,50^2)$. The variance $300^2$ and $50^2$ are chosen based on the modes of our prior distributions for $\sigma^2_r$ and $\sigma^2_\epsilon$ defined in Section 4.2 prior part. After superimposing the synthetic errors, the locations with negative thickness values or $I(s_j) = 0$ are set to zero thicknesses.

Out of the different possible random realizations from the assumed discrepancy model, we choose to use the one shown in Figure 3, which induces decreased ice thickness and less ice coverage in the high discrepancy area compared to the original model output at the assumed truth, to create a challenging example problem (Note that there is one inland location with removed ice due to a random realization from the Bernoulli distribution in step (a)).

Fig 3: The map of original model output for the assumed truth (left), the synthetic observation (middle), and the synthetic errors generated as described in Section 5.4 (right). The generated discrepancy terms remove some of the ice covered regions and reduce the ice thickness in the high discrepancy area, covered by the discrepancy matrices $K_r$ and $K_v$ described in Figure S5.

For both the full and the binary-only approaches, we respectively have obtained an MCMC chain with a length of 1,000,000 iterations and verified that it has reached equilibrium by comparing the first half and the whole MCMC chain (results not shown). The overall computing time took 73 hours on a high-performance single core with an R code implementation. Switching to a faster program language and applying parallelization will certainly make the computation much faster, but we did not seek to improve the computational time here because the application problem at hand does not require a faster solution. To verify the performance of our method in terms recovering the assumed true input parameter setting we compare the estimated posterior densities with the assumed true input parameter settings.

CRHASE, GEO, LITH, and LAPSE are expected to affect the ice thickness over areas already covered with ice, rather than the horizontal extent of the grounded ice sheet. LAPSE
controls the lapse rate and hence the vertical atmospheric temperature profile; the higher the
value of LAPSE, the colder the surface air temperature becomes for a thicker ice sheet, thus
affecting ice thickness through precipitation and surface melt. The other three, CRHASE,
GEO, and LITH, affect interactions between the base of the ice sheet and the bedrock.
CRHASE is a multiplicative factor for the basal sliding factor (CRH) applied only to the
inshore area of the Amundsen Sea Embayment, controlling slipperiness of the bedrock in
that region. GEO sets the geothermal heat flux that flows from the bedrock into the base
of the ice sheet, and hence the amount of basal melting of grounded ice. Lastly, LITH is the
flexural rigidity of the lithosphere in the bedrock model, influencing the vertical deformation
of the bedrock in response to the ice load on it, and hence the elevation of the ice surface.

The resulting posterior densities based on the binary patterns only and the full thickness
patterns are shown in Figures S7 and S8 respectively. For the parameters listed above,
utilizing the full thickness pattern improves the parameter estimation results. The posterior
densities for GEO and LAPSE are notably improved, moving the area with the highest
posterior density closer to their true values. We see a similar effect for LITH as well, but
the peak of the posterior density is still away from the true value after the improvement. The
posterior for CRHASE is improved as well, as the originally severely biased posterior based
on the binary pattern has become less biased, assigning more probability mass towards the
true value with the full thickness pattern. Since glaciologically LAPSE, CRHASE, LITH,
am are expected to be improved when the thickness information is utilized and there are also
notable changes in the posterior densities based on the real observation (see Figures S9
and S10), we conclude that the inference results on these parameters are improved in this
simulated example.

By combining information from the MCMC sample for the input parameters and the
future projection part of the existing ensemble runs, we can generate predictive distributions
of ice volume change. To be more specific, the procedure for generating the predictive
distributions for the future projections consists of the following three steps: (i) building
an emulator using the existing 499 model runs (which are projected into futures based on
future climate scenarios as described in Section 2) while treating input parameters as input
variables and the future ice volume changes as the output variable for the emulator, and
(ii) generating a sample for the predictive distribution of future ice volume changes by
supplying the MCMC sample for input parameters to the emulator built in (i); (iii) finally,
finding the predictive distribution using kernel density estimation based on the predictive
sample obtained in (ii).

We apply the above procedure to our MCMC samples used in Figures S9 and S10 to find
the predictive distributions of the future ice change volumes caused by WAIS ice volume
loss calculated at 500 years from the present. One may consider an emulator for the full
ice volume change trajectories over time as well, but this may require an emulator that can
account for the changing input-output relationship over time, i.e. possible non-stationarity
in the covariance structure for the GP emulator. We consider this direction as a possible
future work. The resulting projections are shown in Figures 4.

As another benchmark, we also generate future projections based on the future ice volume
change values for the original 499 ensemble members, by converting the existing input
parameter settings \( \theta_1, \ldots, \theta_n \) into future ice volume change values using the emulator from
step (ii) above and applying kernel density estimation to them. The ice volume change
projection based on this ‘no calibration’ shown in 4 spans a broad range covering both large increase and decrease in WAIS volume (see Section 5.5 for more detailed discussion). While both the binary only and the full thickness-based calibration results lead to significant improvement from the no calibration result, utilizing the full thickness patterns leads to a sharper predictive distribution with less uncertainty. The peak of the predictive density, however, has not changed. This is perhaps because the negative bias in GEO (which will bias the prediction towards lower values) was compensated by the positive bias in LAPSE (which will bias the prediction towards higher values) in the binary-only result, and hence improving on the modes of these parameters has not led to reduction in bias in the future projection. We see a similar change of the future projection (reduced uncertainty with the same peak location) in the real data result in Section 5.5 as well. The improvement with the full thickness pattern has an important scientific meaning, because the lower tail in the predictive distribution based only on the binary patterns is removed, ruling out the possibility of a large ice volume growth of WAIS.

5.5. Calibration Using Real Observational Data. We now apply our calibration approach to the Bedmap2 dataset introduced in Section 2. To see if the combined basis $\mathbf{K}_+ = [\mathbf{K}_{+,u}, \mathbf{K}_r]$ used here is adequate to represent the observed thickness pattern $\mathbf{Z}^+$ (i.e., we are not in a ‘terminal case’ defined in (Salter et al., 2019) where $\mathbf{K}_+$ cannot accurately capture the spatial pattern in $\mathbf{Z}^+$) we computed the root mean squared errors between $\mathbf{Z}^+$ and its projected pattern onto the basis $\mathbf{K}_+(\mathbf{K}_+^T \mathbf{K}_+)^{-1} \mathbf{K}_+^T \mathbf{Z}^+$. The computed MAE is 29.549 m, which is much smaller than the MAE between $\mathbf{Z}^+$ and the most similar model output (62.119 m).

Unlike the Arctic ice sheet in Greenland, the ice volume change in the West Antarctic region has been considered to be largely uncertain due to the competing effects of increased temperature. On one hand, a warmer climate in the region leads to more precipitation (in the form of snow), which can increase the ice volume of WAIS. On the other hand, a warmer climate will increase the ocean temperature surrounding WAIS and hence can lead to ice volume loss as most of the ice body in the area are marine-based, i.e., its bottom is below the sea level (see Figure 1e). Therefore, our main question of scientific interest is whether utilization of the full ice thickness in calibration, in combination with the glaciologically-motivated data-model discrepancy term, can reduce this uncertainty in parameter estimation and as a result the uncertainty in future projections.

The parameter estimation results based on the real observational data are shown in Figures S9 (only with binary patterns) and S10 (with the full thickness patterns). The density for CRHASE has been shifted towards higher values and hence the information from full thickness pattern supports more slippery bedrock in the ASE region compared to the binary patterns. The peak and the lower tail of the density of GEO has been shifted towards higher values as well, indicating that the full thickness pattern supports higher values of geothermal heat flux than the binary pattern. We also see a similar effect on LAPSE, but towards smaller values, meaning that the information from the full thickness pattern makes a lower lapse rate more plausible. All of these results make increases in WAIS ice volume (i.e., decreases in sea level equivalent (sle)) much less plausible, which is reflected as a much thinner lower tail in the predictive distribution for the future volume change in Figure 5 (projections for 500 years from present).
The peak of the posterior density of LITH, on the other hand, has notably shifted towards higher values, which supports a more rigid lithospheric bedrock component (reducing horizontal contrasts in the bedrock response to the ice load and hence generally less vertical movement). Both LITH and TAU affect bedrock deformation in different ways, and their relative effects depend on whether the ice sheet is in retreat or expansion. Here we can isolate the effect of LITH due to the other constraints, because the estimated posterior for TAU is largely unchanged between the binary only and the full thickness results due to the prior on it, and the predictive distribution of ice volume change puts most of its mass on the retreating ice sheet.

The smaller value of LITH limits the impact of the lower LAPSE estimate based on the full thickness pattern. It also limits the impact of the strong oceanic melting effect due to the ocean temperature increase, which is set to be high by the prior on OFAC in both the binary only calibration and the full thickness calibration. These limiting effects occur because higher values of LITH reduce small-scale responses of the bedrock near the edge of the grounded ice sheet, generally increasing ice surface elevations near the grounding line—this in turn limits the effect of the lower lapse rate as ice-sheet surfaces will be higher, and also limits the effect of ocean melting as a larger portion of the ice sheet will be grounded and not susceptible to oceanic melting. This is reflected in the predictive distribution in Figures 5 as the slightly lower upper limit for projection based on the full thickness. It seems that the compensation between LAPSE and LITH also keeps the peak of the predictive distributions unchanged.

Using the covariance matrix computed based on the posterior samples of the relevant parameters, we compute the test statistics and the corresponding p-values proposed by Salter et al. (2019) to confirm that we are not in the terminal case with the calibration based on \( Z^+ \). To reduce computational cost, we compute the test statistic for every 1,000th MCMC iteration (and after 20,000th iteration to allow for some burn-in). The result is summarized in Figure S11. Most of the computed test statistic values are well below the critical value of 1650.941 based on the Chi-square distribution with the degrees of freedom \( m = 1,558 \) (the number of elements in \( Z^+ \)). Therefore, the basis used in our calibration model can represent the observational data well and we are not in the terminal case here.

6. Summary and Future Directions. In this paper we have proposed an efficient emulation and calibration method for the PSU-3D ice model for the West Atlantic Ice Sheet. Our approach can handle semi-continuous spatial model output and observational data, which often arise in scientific fields such as glaciology and meteorology. The methodology we have described here can also be applied to a wide range of calibration problems that involve semi-continuous spatial or image data. In the field of climate science and meteorology, for example, many important processes such as precipitation, pollution, and storm surge level are in the form of semi-continuous spatial data.

We use a mixture model for the semi-continuous output which results in a multiplicative representation of the likelihood between the binary and continuous part of the dataset. Using dimension reduction and basis representation techniques, our approach can overcome the inferential and computational challenges posed by high-dimensional and dependent semi-continuous data and provide a statistically sound way to quantify input parameter uncertainties. In a simulation setting, we have shown that our approach can recover the true
input parameter values and lead to smaller parametric and prediction uncertainties when compared to methods that simplify the observations and model output, by converting the semi-continuous data into binary data. In our real data application with the Bedmap2 dataset we observe reduction in parametric and prediction uncertainties that are similar to the simulated data example. Through these, we have demonstrated the value of our approach in the context of a well known model for the Antarctic ice sheet.

Possible extensions of the proposed approach are as follows: First, our approach can be easily modified and applied to an application problem that involves model output and observational data in the form of zero-inflated count spatial data. Such data often arise in ecology applications, where the subjects of study such as animal or plant species show zero prevalence in a large portion of the study area. Second, our approach models the binary patterns indirectly through the logit. This forces to define a specific type of 'nugget' effect defined by the marginal Bernoulli distribution at each location. Relaxing this assumption will lead to a more flexible model specification. Finally, another future direction is to formulate a method that can handle tens or more number of input parameters, which may be useful for ice models with a larger number of uncertain parameters. One possibility is to use the sufficient dimension reduction (e.g., Cook and Ni, 2005) or deep neural network (e.g., Goodfellow et al., 2016) that can extract low dimensional features from a large number of predictors that are most relevant to the response variable.

Acknowledgement. The authors are grateful to Dr. David Pollard for the helpful discussion. This material was based upon work partially supported by the National Science Foundation under Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute (SAMSI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

APPENDIX A: MATRIX COMPUTATION IN SECTION 4.3

Let $K_+ = [K_{+u} K_r]$, then the covariance matrix in (4.8) can be rewritten as

$$
\Sigma_+ = [K_{+u} K_r] \Sigma_{\xi,r} [K_{+u} K_r]^T + \sigma_\epsilon^2 I_m.
$$

By applying the Sherman-Morrison-Woodbury formula (Woodbury, 1950), the inverse of this matrix can be expressed as

$$
(K_+ \Sigma_{\xi,r} K_+^T + \sigma_\epsilon^2 I_m)^{-1} = \sigma_\epsilon^{-2} I_m - \sigma_\epsilon^{-2} K_+ \left( \Sigma_{\xi,r}^{-1} + \sigma_\epsilon^{-2} K_+^T K_+ \right)^{-1} K_+^T \sigma_\epsilon^{-2}.
$$

This reduces the order of the computational cost of matrix inversion from $O(n^3)$ to $O(n^2)$.

In a similar fashion, by applying the determinant formula (Harville, 2008) the determinant of the matrix can be rewritten as

$$
\left| K_+ \Sigma_{\xi,r} K_+^T + \sigma_\epsilon^2 I_m \right| = \sigma_\epsilon^{2m} \left| \Sigma_{\xi,r}^{-1} + \sigma_\epsilon^{-2} K_+^T K_+ \right| \left| \Sigma_{\xi,r} \right|
$$

This gives a similar computational gain as the Sherman-Morrison-Woodbury formula.
REFERENCES


Fig 4: Ice volume change projections in sea level change equivalence (sle) based on the estimated posterior densities shown in Figures S7 (dashed and dotted line), S8 (solid line), and the original ensemble with 499 runs (i.e. no calibration applied, shown as dashed grey line). The vertical line shows the projected change for the assumed truth. The curves show the predictive densities and the bars above them show the 95% highest density intervals. The projection based on the full thickness patterns (solid line) has a sharper density with a lower bias than that based on the binary patterns only (dashed and dotted line).
Fig 5: The same as Figure 4 except that the results are based on the densities in Figures S9 and S10, the posterior densities for observational data. Utilizing the full thickness pattern in calibration leads to reduced uncertainty in future projections compared to the result based only on the binary pattern. In particular, it rules out the possibility of large ice volume increase, which is corresponding to the lower tail of the binary only projection.