The Annals of Applied Statistics
2022, Vol. 16, No. 3, 1937–1961
https://doi.org/10.1214/21-AOAS1577

ICE MODEL CALIBRATION USING SEMICONTINUOUS SPATIAL DATA

BY WON CHANG[1,a], BLEDAR A. KONOMI[1,b], GEORGIOS KARAGIANNIS[2,c], YAWEN GUAN[3,d] AND MURALI HARAN[4,e]

[1]*Division of Statistics and Data Science, University of Cincinnati,* [a]*won.chang@uc.edu,* [b]*konomibr@ucmail.uc.edu*
[2]*Department of Mathematical Sciences, Durham University,* [c]*georgios.karagiannis@durham.ac.uk*
[3]*Department of Statistics, University of Nebraska,* [d]*yawen.guan@unl.edu*
[4]*Department of Statistics, Pennsylvania State University,* [e]*mharan@stat.psu.edu*

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 semicontinuous 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) provide a wavelet-based approach for calibration with functional model output. Sansó and Forest (2009) calibrate a climate model with multivariate output while Higdon et al. (2008) and Chang et al. (2014) provide approaches for calibrating models with high-dimensional

Received November 2020; revised October 2021.
Key words and phrases. Computer model emulation, computer model calibration, semicontinuous spatial data, West Antarctic ice sheet, dimension reduction.

spatial data output. More recently, Chang et al. (2016a) develops an approach for high-dimensional binary spatial data output, and Sung et al. (2020) proposes a method for binary time series output. Cao et al. (2018) provide a method for censored functional data. Ice sheet thickness data, including the West Antarctic ice sheet data set we analyze here, are frequently in the form of high-dimensional semicontinuous 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, 2016b), 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 semicontinuous 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 semicontinuous data with truncation at zero poses nontrivial 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 semicontinuous 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 semicontinuous spatial data.

Existing methods for handling semicontinuous data in the spatial statistics literature are based on the truncated Gaussian process approach (Stein (1992), De Oliveira (2005)). In this framework the semicontinuous 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 nonzero 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 approach. 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 semicontinuous 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 semicontinuous 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 semicontinuous 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, DeConto and Alley (2015), Pollard et al. (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 prespecified design points in the input parameter space (often called a perturbed physics ensemble) and then build a statistical surrogate based on these 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 10 varied input parameters. Note that a part of the simulation results was 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 to propose a statistical approach that can properly handle the inferential challenges arising from the data type of the thickness patterns, semicontinuous spatial data with many zeros.

We refer to Table S1 in the Supplementary Material (Chang et al. (2022)) for a detailed description of these 10 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 Supplementary Material.) The ranges of these parameters are determined based on the previous literature (Pollard and DeConto (2012), Pollard, DeConto and Alley (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 the future for 5000 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, Payne and Vieli (2010)) that is uniformly perturbed based on a deep-sea core δ^{18}O record (Pollard and DeConto (2009), Pollard and DeConto (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 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

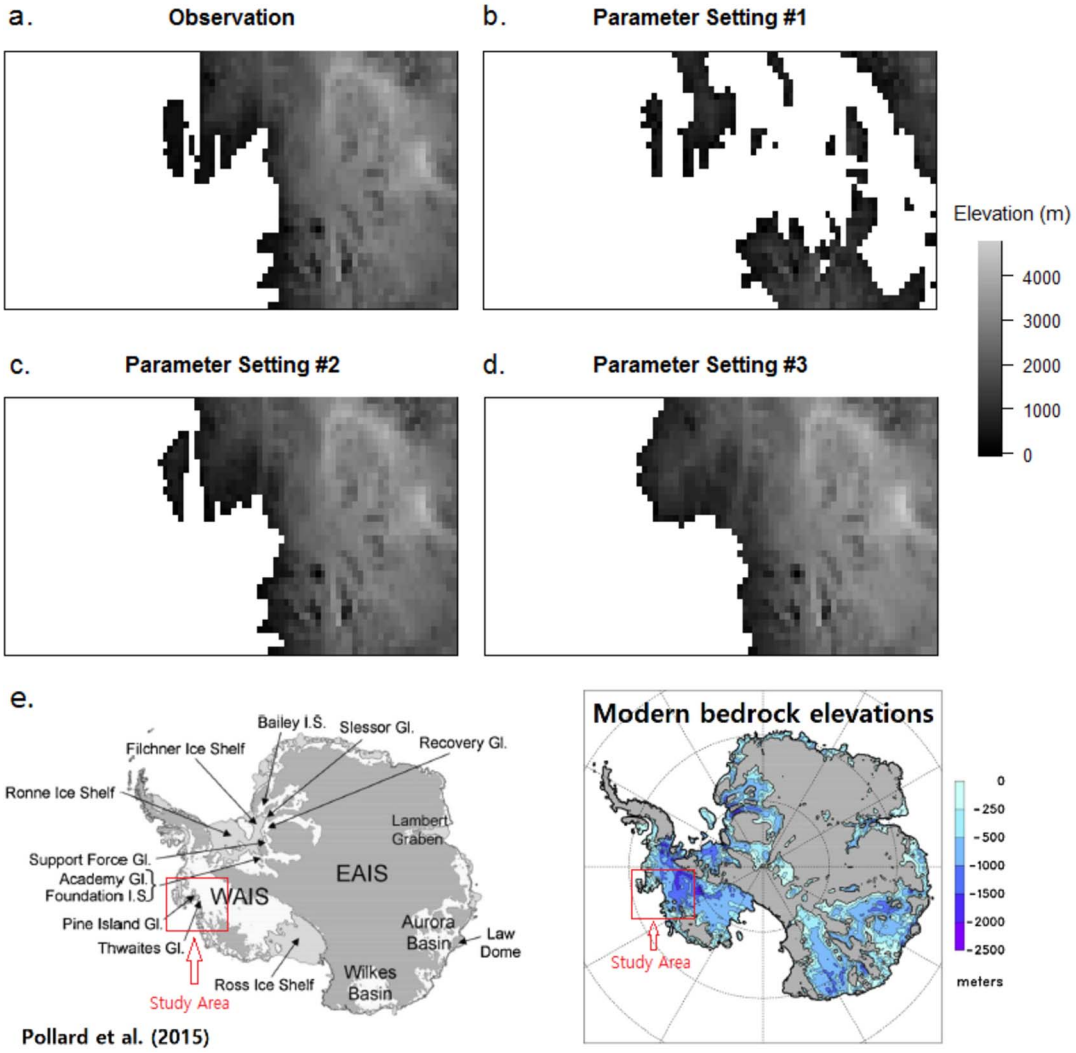

FIG. 1. *Observational data* (*a*) *from Bedmap* 2 *data* (*Fretwell et al.* (2013)) *and example model runs* (*b*)–(*d*) *from PSU-3D ice model*. (*e*) *The maps of region names* (*left*) *and bedrock elevations* (*right*) *that are adopted from Pollard, DeConto and Alley* (2015) *with a rectangular box showing the study area. The ice sheet in the study area is mostly marine-based, that is, its bottom is beneath the sea level.*

region has 86 × 37 pixels with 20 km × 20 km resolution (Figure 1(b)–(d)). Figure 1(e) (adapted from Pollard, DeConto and Alley (2015)) shows the location of the study region in WAIS and the topography of the bedrock in the area.

To calibrate the 10 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 1(a)). 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 averaging 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 semicontinuous spatial data which poses nontrivial statistical challenges for our calibration framework.

3. Computer model emulation and calibration using semicontinuous 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 semicontinuous data. A flowchart describing the overall modeling framework can be found in Figure S1.

We use the following notation hereafter: Let the p-dimensional vector $\mathbf{Y}(\boldsymbol{\theta}) = [Y(\boldsymbol{\theta}, \mathbf{s}_1), \ldots, Y(\boldsymbol{\theta}, \mathbf{s}_p)]^T$ denote the spatial pattern of ice thickness at the spatial locations of the model grid $\mathbf{s}_1, \ldots, \mathbf{s}_p \in R^2$ which is generated from the computer model given input parameter setting $\boldsymbol{\theta} \in R^d$. Here, d is the dimension of the input space which in our application is equal to 10. The observed data at the same spatial locations are denoted as a p-dimensional vector $\mathbf{Z} = [Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_p)]^T$. Here, $Y(\boldsymbol{\theta}, \mathbf{s}_j)$ and $\mathbf{Z}(\mathbf{s}_j)$ can have either positive values representing the ice thickness or zero values denoting the absence of ice at location $\mathbf{s}_j$ (see Figure 1).

We denote the design points for the input parameters in our ensemble as $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$. As a result, the collection of model output in our ensemble can be denoted as an $n \times p$ matrix $\mathbf{Y}$, with elements $[\mathbf{Y}]_{i,j} = Y(\boldsymbol{\theta}_i, \mathbf{s}_j)$ 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 spatial locations. In our ice thickness application the number of spatial locations for the grid is $p = 86 \times 37 = 3182$, and the number of model runs in the ensemble is $n = 499$.

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

1. Create a $n \times p$ matrix for ice-no ice binary patterns, $\{I_y(\boldsymbol{\theta}_i, \mathbf{s}_j)\}(i = 1 \ldots, n, j = 1, \ldots, p)$, by dichotomizing the elements in the model output matrix $\mathbf{Y}$ into 0s and 1s. Apply logistic principal component analysis (LPCA) to the dichotomized output matrix to find the $n \times J_w$ matrix for LPC scores $\mathbf{W}$.
2. Apply likelihood-based PCA only to the nonzero values in $\mathbf{Y}$ to find the $n \times J_u$ matrix for PC scores $\mathbf{U}$.
3. For each column in $\mathbf{W}$ and $\mathbf{U}$, separately construct a one-dimensional GP emulator by finding Bayesian estimates for the emulator parameters with the procedure proposed by Gu, Palomo and Berger (2019). Let $\boldsymbol{\psi}(\boldsymbol{\theta})$ and $\boldsymbol{\xi}(\boldsymbol{\theta})$, respectively, denote J_w- and J_u-dimensional emulated processes for the unobserved values of $\mathbf{w}(\boldsymbol{\theta})$ and $\mathbf{u}(\boldsymbol{\theta})$ which are collections of independently constructed one-dimensional GP emulators.
4. Infer the best input parameter setting $\boldsymbol{\theta}^*$ along with other parameters based on the posterior density, given the observational data $\mathbf{Z}$ (see equation (4.7) for its definition). The Bayesian inference can be facilitated through Metropolis-within-Gibbs sampling.

A flowchart describing the overall modeling framework is included as Figure S1 (Chang et al. (2022)).

3.2. *Computer model emulation using semicontinuous spatial data.* Since only a limited 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 semicontinuous nature of $\mathbf{Y}(\boldsymbol{\theta})$ 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 semicontinuous $Y(\boldsymbol{\theta}, \mathbf{s}_j)$ variable possible, we introduce an indicator variable $I_y(\boldsymbol{\theta}_i, \mathbf{s}_j)$ whose value is one if grounded ice is present at the given parameter setting and spatial location $(\boldsymbol{\theta}_i, \mathbf{s}_j)$ or zero otherwise for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Given that grounded ice is present, we model

the thickness as $Y(\boldsymbol{\theta}_i, \mathbf{s}_j) = q(h(\boldsymbol{\theta}_i, \mathbf{s}_j))$, where $q : \mathbb{R} \to \mathbb{R}_+$ is a bijective transformation function that allows $h(\boldsymbol{\theta}_i, \mathbf{s}_j)$ 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 $Y(\boldsymbol{\theta}_i, \mathbf{s}_j)$ as

$$Y(\boldsymbol{\theta}_i, \mathbf{s}_j) = \begin{cases} q\big(h(\boldsymbol{\theta}_i, \mathbf{s}_j)\big) & \text{if } I_y(\boldsymbol{\theta}_i, \mathbf{s}_j) = 1, \\ 0 & \text{if } I_y(\boldsymbol{\theta}_i, \mathbf{s}_j) = 0, \end{cases} \tag{3.1}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Using this representation, we can translate the problem of emulating $\mathbf{Y}(\boldsymbol{\theta})$ into the problem of finding the predictive distributions of the binary response $I_y(\boldsymbol{\theta}, \mathbf{s}_1), \ldots, I_y(\boldsymbol{\theta}, \mathbf{s}_p)$ and the transformed thickness values $\mathbf{h}(\boldsymbol{\theta}) = [h(\boldsymbol{\theta}, \mathbf{s}_1), \ldots, h(\boldsymbol{\theta}, \mathbf{s}_p)]^T$ at any untried input parameter setting $\boldsymbol{\theta}$. Therefore, we can model $\mathbf{h}(\boldsymbol{\theta})$ directly using an existing method for continuous data, such as basis representation (see, e.g., Chang et al. (2014), Higdon et al. (2008)), since its elements are unbounded and continuous. We use a p-dimensional vector $\boldsymbol{\eta}(\boldsymbol{\theta}) = [\eta(\boldsymbol{\theta}, \mathbf{s}_1), \ldots, \eta(\boldsymbol{\theta}, \mathbf{s}_p)]^T$ to denote the emulated process for $\mathbf{h}(\boldsymbol{\theta})$. (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 $I_y(\boldsymbol{\theta}, \mathbf{s}_1), \ldots, I_y(\boldsymbol{\theta}, \mathbf{s}_p)$ through their corresponding logits $\boldsymbol{\gamma}(\boldsymbol{\theta}) = [\gamma(\boldsymbol{\theta}, \mathbf{s}_1), \ldots, \gamma(\boldsymbol{\theta}, \mathbf{s}_p)]^T$, defined as

$$\gamma(\boldsymbol{\theta}, \mathbf{s}_j) = \log \frac{P(I_y(\boldsymbol{\theta}, \mathbf{s}_j) = 1)}{1 - P(I_y(\boldsymbol{\theta}, \mathbf{s}_j) = 1)}$$

for $j = 1, \ldots, p$, as in Chang et al. (2016a). Since $\boldsymbol{\gamma}(\boldsymbol{\theta}) = [\gamma(\boldsymbol{\theta}, \mathbf{s}_1) \ , \ldots, \gamma(\boldsymbol{\theta}, \mathbf{s}_p)]^T$ can be again treated as continuous variables with unbounded support, an emulation approach for continuous variables can be applicable. Since $\gamma(\boldsymbol{\theta}, \mathbf{s}_j)$ is an unobserved latent variable, even if $\boldsymbol{\theta}$ is one of the existing design points $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$, 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 $\boldsymbol{\eta}(\boldsymbol{\theta})$ and $\boldsymbol{\gamma}(\boldsymbol{\theta})$ at any untried settings $\boldsymbol{\theta}$ (which are possibly dependent on each other).

3.3. *Computer model calibration using semicontinuous spatial data.* In addition to the new emulation framework described above, we formulate a new calibration framework for semicontinuous 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 $Z(\mathbf{s}_j)$, as in (3.1). We define the variable $I_z(\mathbf{s}_j)$ to be an indicator with a value of one if observed grounded ice presents at $\mathbf{s}_j$ and zero otherwise. To transform the observational data, we use the same transformation function q as in (3.1). At any spatial location $\mathbf{s}_j$, we assume observation of ice thickness $Z(\mathbf{s}_j)$ can be represented as follows:

$$Z(\mathbf{s}_j) = \begin{cases} q\big(t(\mathbf{s}_j)\big) & \text{if } I_z(\mathbf{s}_j) = 1, \\ 0 & \text{if } I_z(\mathbf{s}_j) = 0. \end{cases} \tag{3.2}$$

In a similar fashion to our emulation framework, we set up our model for the transformed thickness $t(\mathbf{s}_j)$ and the logit of $I_z(\mathbf{s}_j)$, denoted as $\lambda(\mathbf{s}_j)$. 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 $(\gamma(\boldsymbol{\theta}^*, \mathbf{s}_j))$ while accounting for data-model discrepancy:

$$\lambda(\mathbf{s}_j) = \gamma\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big) + \alpha(\mathbf{s}_j), \tag{3.3}$$

where $\boldsymbol{\theta}^*$ is the input parameter setting that gives the "best" match between model output and observational data and $\alpha(\mathbf{s}_i)$ is a spatially correlated discrepancy term, normally distributed, representing the sum of spatially correlated parts of the model structural error (i.e., model misrepresentation 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 $\mathbf{s}_j$ given the logit $\lambda(\mathbf{s}_j)$, as in a usual generalized linear model setting.

The model for $t(\mathbf{s}_j)$ needs to be defined only for the locations with $I_z(\mathbf{s}_j) = 1$. Let $m = \sum_{j=1}^{p} I_z(\mathbf{s}_j)$ 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, $\mathbf{Z}^+ = [Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_m)]$, are positive while the rest, $Z(\mathbf{s}_{m+1}), \ldots, Z(\mathbf{s}_p)$, are 0. For $\mathbf{s}_1, \ldots, \mathbf{s}_m$, we use the following model for the transformed thickness:

$$t(\mathbf{s}_j) = \eta\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big) + \delta(\mathbf{s}_j) + \epsilon(\mathbf{s}_j), \tag{3.4}$$

where the random variables $\boldsymbol{\delta} = [\delta(\mathbf{s}_1), \ldots, \delta(\mathbf{s}_m)]^T \sim N(\mathbf{0}, \Sigma_\delta)$ and $\boldsymbol{\epsilon} = [\epsilon(\mathbf{s}_1), \ldots, \epsilon(\mathbf{s}_m)]^T \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_m)$, 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 Σ_δ reflects the spatial dependence among $\delta(\mathbf{s}_1), \ldots, \delta(\mathbf{s}_m)$.

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 $t(\mathbf{s})$ and for the calibration model $\lambda(\mathbf{s})$ have the same interpretation; this allows the emulator for $h(\boldsymbol{\theta}, \mathbf{s})$ to be directly used in modeling $\lambda(\mathbf{s})$.

The model in (3.3) assigns the following Bernoulli distribution for $I_z(\mathbf{s}_j)$ (conditionally on the value of $\boldsymbol{\theta}^*$ and discrepancy $\alpha(\mathbf{s}_j)$):

$$\begin{aligned} &P\big(I_z(\mathbf{s}_j) = x | \gamma\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big), \alpha(\mathbf{s}_j)\big) \\ &\quad = \left(\frac{\exp(\gamma(\boldsymbol{\theta}^*, \mathbf{s}_j) + \alpha(\mathbf{s}_j))}{1 + \exp(\gamma(\boldsymbol{\theta}^*, \mathbf{s}_j) + \alpha(\mathbf{s}_j))}\right)^x \left(\frac{1}{1 + \exp(\gamma(\boldsymbol{\theta}^*, \mathbf{s}_j) + \alpha(\mathbf{s}_j))}\right)^{1-x}. \end{aligned}$$

Given this distribution for $I_z(\mathbf{s}_j)$, we can view the specification in (3.2) as a mixture model with density

$$\begin{aligned} &f\big(Z(\mathbf{s}_j) | \eta\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big), \delta(\mathbf{s}_j), \sigma_\epsilon^2, \gamma\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big), \alpha(\mathbf{s}_j)\big) \\ &\quad = \left|\frac{\partial t(\mathbf{s}_j)}{\partial Z(\mathbf{s}_j)}\right| f\big(t(\mathbf{s}_j) | \eta\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big), \delta(\mathbf{s}_j), \sigma_\epsilon^2\big) P\big(I_z(\mathbf{s}_j) = 1 | \gamma\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big), \alpha(\mathbf{s}_j)\big) \\ &\qquad + \mathcal{D}_0\big(Z(\mathbf{s}_j)\big) P\big(I_z(\mathbf{s}_j) = 0 | \gamma\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big), \alpha(\mathbf{s}_j)\big) \end{aligned} \tag{3.5}$$

for all locations $\mathbf{s}_1, \ldots, \mathbf{s}_p$, where the density function $f(t(\mathbf{s}_j) | \eta(\boldsymbol{\theta}^*, \mathbf{s}_j), \delta(\mathbf{s}_j), \sigma_\epsilon^2)$ is given by (3.4) and $\mathcal{D}_0$ is the Dirac delta function. Since the density in (3.5) can be rewritten as

$$\begin{aligned} &f\big(t(\mathbf{s}_j) | \eta\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big), \delta(\mathbf{s}_j), \sigma_\epsilon^2\big) \\ &\quad = \begin{cases} f\big(t(\mathbf{s}_j) | \eta\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big), \delta(\mathbf{s}_j), \sigma_\epsilon^2\big) P\big(I_z(\mathbf{s}_j) = 1 | \gamma\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big), \alpha(\mathbf{s}_j)\big) & \text{if } I_z(\mathbf{s}_j) = 1, \\ P\big(I_z(\mathbf{s}_j) = 0 | \gamma\big(\boldsymbol{\theta}^*, \mathbf{s}_j\big), \alpha(\mathbf{s}_j)\big) & \text{if } I_z(\mathbf{s}_j) = 0, \end{cases} \end{aligned}$$

and $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_p)$ are conditionally independent, given the relevant parameters, the likelihood for $\mathbf{Z}$ can be factorized as follows:

$$\begin{aligned}
&\mathcal{L}(\mathbf{Z}|\eta^+(\boldsymbol{\theta}^*), \boldsymbol{\delta}, \sigma_\epsilon^2, \boldsymbol{\gamma}(\boldsymbol{\theta}^*), \boldsymbol{\alpha}) \\
&\quad \propto \prod_{j=1}^{m} f\big(t(\mathbf{s}_j)|\eta(\boldsymbol{\theta}^*, \mathbf{s}_j), \delta(\mathbf{s}_j), \sigma_\epsilon^2\big) P\big(I_z(\mathbf{s}_j) = 1|\gamma(\boldsymbol{\theta}^*, \mathbf{s}_j), \alpha(\mathbf{s}_j)\big) \\
&\quad \times \prod_{j=m+1}^{p} P\big(I_z(\mathbf{s}_j) = 0|\gamma(\boldsymbol{\theta}^*, \mathbf{s}_j), \alpha(\mathbf{s}_j)\big) \\
&\quad = \mathcal{L}_1(\mathbf{Z}^+|\eta^+(\boldsymbol{\theta}^*), \boldsymbol{\delta}, \sigma_\epsilon^2)\mathcal{L}_2\big(I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)|\boldsymbol{\gamma}(\boldsymbol{\theta}^*), \boldsymbol{\alpha}\big),
\end{aligned} \tag{3.6}$$

where

$$\mathcal{L}_1(\mathbf{Z}^+|\eta^+(\boldsymbol{\theta}^*), \boldsymbol{\delta}, \sigma_\epsilon^2) = \prod_{j=1}^{m} f\big(t(\mathbf{s}_j)|\eta(\boldsymbol{\theta}^*, \mathbf{s}_j), \delta(\mathbf{s}_j), \sigma_\epsilon^2\big),$$

$$\begin{aligned}
\mathcal{L}_2\big(I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)|\boldsymbol{\gamma}(\boldsymbol{\theta}^*), \boldsymbol{\alpha}\big) = &\prod_{j=1}^{m} P\big(I_z(\mathbf{s}_j) = 1|\gamma(\boldsymbol{\theta}^*, \mathbf{s}_j), \alpha(\mathbf{s}_j)\big) \\
&\times \prod_{j=m+1}^{p} P\big(I_z(\mathbf{s}_j) = 0|\gamma(\boldsymbol{\theta}^*, \mathbf{s}_j), \alpha(\mathbf{s}_j)\big).
\end{aligned}$$

Here, $\eta^+(\boldsymbol{\theta}^*)$ is the vector of emulated process for all positive $Z(\mathbf{s}_j)$'s (i.e., $\eta^+(\boldsymbol{\theta}^*) = [\eta(\boldsymbol{\theta}^*, \mathbf{s}_1), \ldots, \eta(\boldsymbol{\theta}^*, \mathbf{s}_m)]^T$), and $\boldsymbol{\alpha} = [\alpha(\mathbf{s}_1), \ldots, \alpha(\mathbf{s}_p)]^T$. The Jacobian factors $|\frac{\partial t(\mathbf{s}_j)}{\partial Z(\mathbf{s}_j)}|$ 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 $\mathbf{Z}$ can be factored into two parts, one for the positive observations $\mathbf{Z}^+$ and the other for the indicator variables at all locations, $I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)$. This has an important implication for inference on $\boldsymbol{\theta}^*$: utilizing the ice thickness pattern for calibration is essentially using the additional information from the positive ice thickness values $\mathbf{Z}^+$ on top of the binary spatial pattern of ice presence $(I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p))$ in calibration. We will show how this added information improves our inference on the input parameter $\boldsymbol{\theta}^*$ in both the simulated and the real data examples in Section 5 below.

Note that this formulation does not necessarily require independence between $\mathbf{Z}^+$ and $I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)$, because dependence can easily be specified through dependence between $\eta(\boldsymbol{\theta}^*, \mathbf{s}_j)$ and $\gamma(\boldsymbol{\theta}^*, \mathbf{s}_j)$ or $\delta(\mathbf{s}_j)$ and $\alpha(\mathbf{s}_j)$. This is how we impose dependence between $\mathbf{Z}^+$ and $I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)$ 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 $499 \times 3182 \approx 1.6$ 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 ($2p = 6364$), 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 $\boldsymbol{\theta}_i$ and $\mathbf{s}_j$ with $I_y(\boldsymbol{\theta}_i, \mathbf{s}_j) = 1$. In our calibration problem the number of $(\boldsymbol{\theta}_i, \mathbf{s}_j)$ combinations with $I_y(\boldsymbol{\theta}_i, \mathbf{s}_j) = 1$ 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. (2019)).

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 $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}(\boldsymbol{\theta}_1), \ldots, \boldsymbol{\gamma}(\boldsymbol{\theta}_n)]^T$ be a matrix of logits for the binary patterns $\{I_y(\boldsymbol{\theta}_i, \mathbf{s}_j)\}$ $(i = 1, \ldots, n$ and $j = 1, \ldots, p)$ for the existing model runs. The rows of $\boldsymbol{\Gamma}$ correspond to the design points in input parameter settings $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ while the columns are for different spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_p$. We apply logistic principal component analysis (LPCA) (Lee, Huang and Hu (2010)) to decompose the logit matrix $\boldsymbol{\Gamma}$ in the following way:

$$\boldsymbol{\Gamma} = \mathbf{1}_n \boldsymbol{\mu}^T + \mathbf{W}\mathbf{K}_w^T, \tag{4.1}$$

where $\boldsymbol{\mu}$ is the $p \times 1$ mean vector for the spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_p$ (i.e., the column means of $\boldsymbol{\Gamma}$), $\mathbf{W}$ is the $n \times J_w$ logistic principal component (LPC) score matrix, and $\mathbf{K}_w$ is the $p \times J_w$ LPC matrix with a prespecified number of principal components $J_w \geq 1$. The rows of $\mathbf{W} = [\mathbf{w}(\boldsymbol{\theta}_1), \ldots, \mathbf{w}(\boldsymbol{\theta}_n)]^T$ correspond to the logits for different input parameter settings where $\mathbf{w}(\boldsymbol{\theta}) = [w_1(\boldsymbol{\theta}), \ldots w_{J_w}(\boldsymbol{\theta})]^T$ denotes a vector of the LPC scores at $\boldsymbol{\theta}$. The parameters in (4.1) ($\boldsymbol{\mu}$, $\mathbf{W}$, and $\mathbf{K}_w$) can be estimated by maximizing the corresponding likelihood function for these parameters, given the binary patterns $\{I_y(\boldsymbol{\theta}_i, \mathbf{s}_j)\}$ for existing model runs using the minorization and maximization (MM) algorithm. We predict the logits $\boldsymbol{\gamma}(\boldsymbol{\theta})$ at any untried setting $\boldsymbol{\theta}$ by predicting the corresponding LPC scores $\mathbf{w}(\boldsymbol{\theta})$. Each score $w_k(\boldsymbol{\theta})$ (for $k = 1, \ldots J_w$) can be predicted separately using a GP emulator; see Section S3 in the Supplementary Material for details.) We denote the resulting emulated process of LPC scores at $\boldsymbol{\theta}$ as $\boldsymbol{\psi}(\boldsymbol{\theta}) = [\psi_1(\boldsymbol{\theta}), \ldots, \psi_{J_w}(\boldsymbol{\theta})]^T$.

We also apply a likelihood-based PCA method for data with missing values to build an emulator for the ice-thickness patterns. For $\boldsymbol{\theta}_i$ and $\mathbf{s}_j$ with $I_y(\boldsymbol{\theta}_i, \mathbf{s}_j) = 1$, we assume the following model for dimension reduction:

$$h(\boldsymbol{\theta}_i, \mathbf{s}_j) = \sum_{l=1}^{J_u} k_{u,jl} u_l(\boldsymbol{\theta}_i) + e_{ij} \tag{4.2}$$

with $e_{ij} \sim$ 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 $u_l(\boldsymbol{\theta}_i)$ ($i = 1, \ldots, n$ and $l = 1, \ldots, J_u$). Again, $J_u \geq 1$ is the predetermined number of principal components being used for our dimension reduction. This is essentially PCA with missing values, and, therefore, the PC loadings and scores can be estimated via EM algorithm (Stacklies et al. (2007)). We denote the resulting $p \times J_u$ loading matrix by $\mathbf{K}_u$, with (i, j)th element, given by $k_{u,ij}$. In a similar manner to the problem of emulating logits, we predict the latent variables for the thickness $h(\boldsymbol{\theta}, \mathbf{s}_j)$ at any untried setting $\boldsymbol{\theta}$ and location $\mathbf{s}_j$ with a positive thickness value by predicting the corresponding principal component scores $\mathbf{u}(\boldsymbol{\theta}) = [u_1(\boldsymbol{\theta}), \ldots, u_{J_u}(\boldsymbol{\theta})]^T$.

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

$$\operatorname{Cov}\big(u_l(\boldsymbol{\theta}), u_l(\boldsymbol{\theta}')\big) = \zeta_{u,l} I\big(\boldsymbol{\theta} = \boldsymbol{\theta}'\big) + \kappa_{u,l} \exp\left(-\sum_{b=1}^{d} \frac{|\theta_b - \theta_b'|}{\phi_{u,lb}}\right) \tag{4.3}$$

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

$$E\big(u_l(\boldsymbol{\theta}_i) | w_1(\boldsymbol{\theta}_i), \ldots, w_{J_w}(\boldsymbol{\theta}_i)\big) = \sum_{k=1}^{J_w} g_{lk}\big(w_k(\boldsymbol{\theta}_i)\big), \tag{4.4}$$

where the function g_{lk} is given by a natural spline regression model whose degrees of freedom is determined through cross-validation (Hastie (1992)). We let $\boldsymbol{\beta}_{lk}$ be the vector of coefficients for $g_{lk}(\cdot)$ whose dimensionality is the same as the degrees of freedom of g_{lk}. To construct the GP emulator, we find the posterior modes of the covariance parameters (denoted as $\hat{\zeta}_{u,l}$, $\hat{\kappa}_{u,l}$ and $\hat{\phi}_{u,l1}, \ldots, \hat{\phi}_{u,ld}$) and the parameters for the spline functions (denoted as $\hat{\boldsymbol{\beta}}_{l1}, \ldots, \hat{\boldsymbol{\beta}}_{lJ_w}$) for each lth principal component separately using the robust Bayesian procedure (Gu, Wang and Berger (2018), Gu, Palomo and Berger (2019)) as we do for the LPC scores above. When we predict $u_l(\boldsymbol{\theta})$ for any untried setting $\boldsymbol{\theta} \notin \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n\}$, we replace $w_k(\boldsymbol{\theta})$ with $E(\psi_k(\boldsymbol{\theta}) | w_k(\boldsymbol{\theta}_1), \ldots, w_k(\boldsymbol{\theta}_n))$, given by the Gaussian process emulator described above, since $w_k(\boldsymbol{\theta})$ is not available if $\boldsymbol{\theta} \notin \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n\}$. We let $\boldsymbol{\xi}(\boldsymbol{\theta}) = [\xi_1(\boldsymbol{\theta}), \ldots, \xi_{J_u}(\boldsymbol{\theta})]^T$ denote the resulting emulated process for $\mathbf{u}(\boldsymbol{\theta})$.

For any untried input parameter setting $\boldsymbol{\theta}$, we can predict the ice thickness pattern from our computer model in the following two steps: (i) We first predict the logits of ice-no ice patterns $\boldsymbol{\gamma}(\boldsymbol{\theta})$ as $\mathbf{K}_w \boldsymbol{\psi}(\boldsymbol{\theta})$ and (ii) for each location $\mathbf{s}_j$ with $\gamma(\boldsymbol{\theta}, \mathbf{s}_j) > 0$. The predicted thickness is given as $q(\sum_{l=1}^{J_u} k_{u,jl} u_l(\boldsymbol{\theta}))$. Note, however, that the thresholding of the logits at 0 is needed only for evaluating emulation performance (such as generating predicted 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 modes, except for the partial sill parameters for $\boldsymbol{\xi}$, $\boldsymbol{\kappa}_u = [\kappa_{u,1}, \ldots, \kappa_{u,J_u}]$. The partial sill parameters for $\boldsymbol{\xi}$ will be reestimated along other parameters in the calibration model to account for any possible discrepancies in scale (see, e.g., Bhat et al. (2012), Chang et al. (2014, 2015, 2016b), for similar approaches). However, the partial sills for $\boldsymbol{\psi}$ will be fixed at their MLEs without being reestimated in the calibration stage because the binary patterns usually do not have enough information for the scale parameters of the latent variables and hence reestimation for the partial sill parameters often cause identifiability issues, as discussed in Chang et al. (2016a).

4.2. *Calibration using basis representation.* Using the emulators for principal components ($\boldsymbol{\psi}$ and $\boldsymbol{\xi}$) 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 $t(\mathbf{s}_j)$ in (3.4) as

$$t(\mathbf{s}_j) = \sum_{l=1}^{J_u} k_{u,jl}\xi_l(\boldsymbol{\theta}^*) + \sum_{k=1}^{J_r} k_{r,jk} r_k + \epsilon_j \tag{4.5}$$

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

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

$$\boldsymbol{\lambda} = \boldsymbol{\mu} + \mathbf{K}_w \boldsymbol{\psi}(\boldsymbol{\theta}^*) + \mathbf{K}_v \mathbf{v}, \tag{4.6}$$

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

The discrepancy basis matrices $\mathbf{K}_r$ and $\mathbf{K}_v$ 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 $\mathbf{K}_v$ and $\mathbf{K}_r$ are both defined using exponential kernels with knot locations marked as red dots in Figure S5, and their range is defined as 80 km. They are defined based on our prior knowledge that: (i) the used PSU-3D ice model runs with 20 km 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, 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 bedrock 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 500–1000 m, 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 $\sum_{l=1}^{J_u} k_{u,jl}\xi_l(\boldsymbol{\theta}_i)$ 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 $\lambda(\mathbf{s})$. 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 $\mathbf{Y}(\boldsymbol{\theta})$ might be missing in the observational data $\mathbf{Z}$ 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 ($\mathbf{K}_u$ and $\mathbf{K}_w$) accordingly by deleting the rows corresponding to the missing locations.

4.3. *Bayesian inference*. Given the above formulation, we conduct Bayesian inference on $\boldsymbol{\theta}^*$ 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 $\boldsymbol{\theta}^*$, while accounting for other sources of uncertainties despite the complexity of our model specification. All the emulator parameters for $\boldsymbol{\psi}(\boldsymbol{\theta}^*)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}^*)$, except for the sill parameters for $\boldsymbol{\psi}(\boldsymbol{\theta}^*)$ ($\boldsymbol{\kappa}_u$), 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 $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$ now becomes

$$f\big(Z(\mathbf{s}_j)|\boldsymbol{\xi}(\boldsymbol{\theta}^*), \mathbf{r}, \sigma_\epsilon^2, \boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\big) = \left|\frac{\partial t(\mathbf{s}_j)}{\partial Z(\mathbf{s}_j)}\right| f\big(Z(\mathbf{s}_j)|\boldsymbol{\xi}(\boldsymbol{\theta}^*), \mathbf{r}, \sigma_\epsilon^2\big) P\big(I_z(\mathbf{s}_j) = 1|\boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\big) + \mathcal{D}_0\big(Z(\mathbf{s}_j)\big) P\big(I_z(\mathbf{s}_j) = 0|\boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\big)$$

for all locations $\mathbf{s}_1, \ldots, \mathbf{s}_p$, where the density function $f(\cdot|\boldsymbol{\xi}(\boldsymbol{\theta}^*), \mathbf{r}, \sigma_\epsilon^2)$ is for the case with $I_z(\mathbf{s}_j) = 1$ in (4.5). As a result, we have the following likelihood function for $\mathbf{Z}$:

$$\begin{aligned}\mathcal{L}\big(\mathbf{Z}|\boldsymbol{\xi}(\boldsymbol{\theta}^*), \mathbf{r}, \sigma_\epsilon^2, \boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\big) &\propto \prod_{j=1}^{m} f\big(Z(\mathbf{s}_j)|\boldsymbol{\xi}(\boldsymbol{\theta}^*), \mathbf{r}, \sigma_\epsilon^2\big) P\big(I_z(\mathbf{s}_j) = 1|\boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\big) \\ &\quad \times \prod_{j=m+1}^{p} P\big(I_z(\mathbf{s}_j) = 0|\boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\big) \\ &= \mathcal{L}_1\big(\mathbf{Z}^+|\boldsymbol{\xi}(\boldsymbol{\theta}^*), \mathbf{r}, \sigma_\epsilon^2\big)\mathcal{L}_2\big(I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)|\boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\big),\end{aligned}$$

where

$$\begin{aligned}&\mathcal{L}_1\big(\mathbf{Z}^+|\boldsymbol{\xi}(\boldsymbol{\theta}^*), \mathbf{r}, \sigma_\epsilon^2\big) = \prod_{j=1}^{m} f\big(Z(\mathbf{s}_j)|\boldsymbol{\xi}(\boldsymbol{\theta}^*), \mathbf{r}, \sigma_\epsilon^2\big), \\ &\mathcal{L}_2\big(I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)|\boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\big) \\ &\quad = \prod_{j=1}^{m} P\big(I_z(\mathbf{s}_j) = 1|\boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\big) \prod_{j=m+1}^{p} P\big(I_z(\mathbf{s}_j) = 0|\boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\big).\end{aligned}$$

We have a similar factorization, as in (3.6), with one factor for the positive observations $\mathbf{Z}^+$ and the other for the binary variables at all locations $I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)$.

Prior. To complete the Bayesian model specification, we assign the following priors for the model parameters ($\boldsymbol{\theta}^*$, $\mathbf{v}$, σ_r^2, σ_ϵ^2, σ_v^2, $\boldsymbol{\kappa}_u$, and $\mathbf{R}$) in our calibration step:

$$\begin{aligned}&v_j|\sigma_v^2 \sim N(0, \sigma_v^2), \quad j = 1, \ldots, J_v; \qquad \sigma_v^2 \sim IG(2, 1); \\ &\sigma_r^2 \sim IG(a_r, b_r); \qquad \sigma_\epsilon^2 \sim IG(a_\epsilon, b_\epsilon); \\ &\kappa_{u,j} \sim IG(5, 6\hat{\kappa}_{u,j}), \quad j = 1, \ldots, J_u; \qquad \mathbf{R} \sim f(\mathbf{R}); \\ &\boldsymbol{\theta}^* \sim d\pi(\boldsymbol{\theta}^*),\end{aligned}$$

where $f(\mathbf{R})$ is a uniform distribution within the range such that $\mathbf{I}_{J_r} - \mathbf{R}\mathbf{R}^T$ is positive definite, that is, $f(\mathbf{R}) \propto I(\mathbf{I}_{J_r} - \mathbf{R}\mathbf{R}^T$ is positive definite). The shape (a_r) and the scale

(b_r) parameters for the variance of thickness discrepancy σ_r^2 is set to be $a_r = 50$ and $b_r = 500a_r / \max(\mathrm{diag}(\sigma_r^2 \mathbf{K}_r \mathbf{K}_r^T))$, where $\mathrm{diag}(\sigma_r^2 \mathbf{K}_r \mathbf{K}_r^T)$ means the diagonal elements of $\sigma_r^2 \mathbf{K}_r \mathbf{K}_r^T$. This encourages the variance of the resulting discrepancy term $\mathbf{K}_r \mathbf{r}$ (i.e., the diagonal elements of the covariance matrix $\sigma_r^2 \mathbf{K}_r \mathbf{K}_r^T$) to be around 500^2, 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 500–1000 m). The mode of the resulting inverse-Gamma distribution is around 300^2. Similarly, the prior for σ_ϵ^2 is designed to have a mode around 100^2 with $a_\epsilon = 50$ and $b_\epsilon = 100^2(a_\epsilon + 1)$, to reflect our belief that the magnitudes of differences between the model output and the observational data are 50–200 m in the region other than the high-discrepancy area shown in Figure S5, covered by the discrepancy term $\mathbf{K}_r \mathbf{r}$.

As per the prior for $\boldsymbol{\theta}$, we use prior densities inspired by previous calibration results for OCFAC, CALV, CRH, and TAU based on modern and paleodata ("Simple Method" in Figure 2 by Pollard et al. (2016)). The informative priors used here are

$$\mathrm{OCFAC} \sim \mathrm{TN}\big(0.75, 0.1^2, 0, 1\big); \qquad \mathrm{CALV} \sim \mathrm{TN}\big(0.3, 0.15^2, 0, 1\big);$$

$$\mathrm{CRH} \sim \mathrm{TN}\big(0.8, 0.1^2, 0, 1\big); \qquad \mathrm{TAU} \sim \mathrm{TN}\big(0.4, 0.15^2, 0, 1\big),$$

where $\mathrm{TN}(\mu, \sigma^2, \mathrm{low}, \mathrm{upper})$ denotes a truncated normal distribution with the mean parameter, μ, variance, σ^2, and the lower and upper bounds, lower and 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 σ_v^2 with a shape parameter of 2 to avoid possible issues with improper posterior and computational instability (Berger, De Oliveira and Sansó (2001)). We have noticed that the posterior analysis is largely insensitive to the choice of the prior hyperparameters (see Section 5.5 for further discussion on this). For the reestimated 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:

$$\begin{aligned} \pi\big(\boldsymbol{\theta}^*, \mathbf{v}, \sigma_r^2, \sigma_\epsilon^2, \sigma_v^2, \boldsymbol{\kappa}_u, \mathbf{R} | \mathbf{Z}\big) &\propto \pi_1\big(\sigma_r^2, \sigma_\epsilon^2, \boldsymbol{\kappa}_u, \mathbf{R} | \boldsymbol{\theta}^*, \mathbf{v}, \mathbf{Z}^+\big) \\ &\quad \times \pi_2\big(\boldsymbol{\theta}^*, \mathbf{v}, \sigma_v^2 | I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)\big). \end{aligned} \tag{4.7}$$

The first part on the right-hand side is based on the likelihood for $\mathbf{Z}^+$ ($\mathcal{L}_1$) and the relevant priors and obtained by

$$\begin{aligned} \pi_1\big(\sigma_r^2, \sigma_\epsilon^2, \boldsymbol{\kappa}_u, \mathbf{R} | \boldsymbol{\theta}^*, \mathbf{v}, \mathbf{Z}^+\big) &\propto \int\int \mathcal{L}_1\big(\mathbf{Z}^+ | \boldsymbol{\xi}(\boldsymbol{\theta}^*), \mathbf{r}, \sigma_\epsilon^2\big) f\big(\boldsymbol{\xi}(\boldsymbol{\theta}^*) | \boldsymbol{\theta}^*, \boldsymbol{\kappa}_u\big) f\big(\mathbf{r} | \sigma_r^2, \mathbf{v}\big)\, d\mathbf{r}\, d\boldsymbol{\xi} \\ &\quad \times f\big(\sigma_r^2\big) f\big(\boldsymbol{\kappa}_u\big) f\big(\sigma_\epsilon^2\big) f\big(\mathbf{R}\big) \\ &= \mathcal{L}_1^*\big(\mathbf{Z}^+ | \boldsymbol{\theta}^*, \sigma_r^2, \sigma_\epsilon^2, \boldsymbol{\kappa}_u, \mathbf{v}, \mathbf{R}\big) f\big(\sigma_r^2\big) f\big(\boldsymbol{\kappa}_u\big) f\big(\sigma_\epsilon^2\big) f\big(\mathbf{R}\big), \end{aligned}$$

where $f(\sigma_r^2)$, $f(\boldsymbol{\theta}^*)$, $f(\boldsymbol{\kappa}_u)$, $f(\sigma_\epsilon^2)$, and $f(\mathbf{R})$ are the prior densities (defined above) and the marginal likelihood $\mathcal{L}_1^*$ can be written as

$$\begin{aligned}&\mathcal{L}_1^*(\mathbf{Z}^+|\boldsymbol{\theta}^*, \sigma_r^2, \sigma_\epsilon^2, \boldsymbol{\kappa}_u, \mathbf{v}, \mathbf{R})\\&\quad\propto |\Sigma_+|^{-1/2}\exp\biggl[-\frac{1}{2}\bigl(q^{-1}(\mathbf{Z}^+)-\boldsymbol{\mu}_+\bigr)^T\Sigma_+^{-1}(\boldsymbol{\theta}^*)\bigl(q^{-1}(\mathbf{Z}^+)-\boldsymbol{\mu}_+\bigr)\biggr],\end{aligned}$$

with $q^{-1}(\mathbf{Z}^+) = [q^{-1}(Z(\mathbf{s}_1)), \ldots, q^{-1}(Z(\mathbf{s}_m))]^T$. The mean and covariance of $q^{-1}(\mathbf{Z}^+)$ are given by

$$\begin{aligned}\boldsymbol{\mu}_+ &= \mathbf{K}_{+,u}\boldsymbol{\mu}_\xi(\boldsymbol{\theta}^*) - \mathbf{K}_r\boldsymbol{\mu}_{\mathbf{r}|\mathbf{v}},\\ \Sigma_+ &= [\mathbf{K}_{+,u}\mathbf{K}_r]\Sigma_{\xi,r}[\mathbf{K}_{+,u}\mathbf{K}_r]^T + \sigma_\epsilon^2\mathbf{I}_m.\end{aligned} \tag{4.8}$$

Here, $\boldsymbol{\mu}_\xi(\boldsymbol{\theta}^*)$ is the mean of the emulated process $\boldsymbol{\xi}(\boldsymbol{\theta}^*)$, 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

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

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

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{r}|\mathbf{v}} &= \frac{\sigma_r}{\sigma_v}\mathbf{R}\mathbf{v},\\ \Sigma_{\mathbf{r}|\mathbf{v}} &= \sigma_r^2(\mathbf{I}_{J_r} - \mathbf{R}\mathbf{R}^T).\end{aligned}$$

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 the Appendix for further details.

The second part of the posterior density is given as

$$\begin{aligned}&\pi_2(\boldsymbol{\theta}^*, \mathbf{v}, \sigma_v^2|I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p))\\&\quad\propto \mathcal{L}_2\bigl(I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)|\boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\bigr) f\bigl(\boldsymbol{\psi}(\boldsymbol{\theta}^*)|\boldsymbol{\theta}^*\bigr) f(\boldsymbol{\theta}^*) f\bigl(\mathbf{v}|\sigma_v^2\bigr) f(\sigma_v^2)\end{aligned}$$

with the prior densities $f(\boldsymbol{\psi}(\boldsymbol{\theta}^*)|\boldsymbol{\theta}^*)$, $f(\boldsymbol{\theta}^*)$, $f(\mathbf{v}|\sigma_v^2)$, and $f(\sigma_v^2)$. The formulation for the logits in (4.6) leads to the following likelihood function for $I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)$:

$$\mathcal{L}_2\bigl(I_z(\mathbf{s}_1), \ldots, I_z(\mathbf{s}_p)|\boldsymbol{\psi}(\boldsymbol{\theta}^*), \mathbf{v}\bigr) \propto \prod_{j=1}^p\biggl[\biggl(\frac{\exp(\lambda(\mathbf{s}_j))}{1+\exp(\lambda(\mathbf{s}_j))}\biggr)^{I_z(\mathbf{s}_j)}\biggl(\frac{1}{1+\exp(\lambda(\mathbf{s}_j))}\biggr)^{1-I_z(\mathbf{s}_j)}\biggr],$$

where the logits $\lambda(\mathbf{s}_1), \ldots, \lambda(\mathbf{s}_p)$ are determined by $\boldsymbol{\psi}(\boldsymbol{\theta}^*)$ and $\mathbf{v}$ through the basis representation in (4.6). The prior for $\boldsymbol{\psi}(\boldsymbol{\theta}^*)$ 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:

$$f\bigl(\boldsymbol{\psi}(\boldsymbol{\theta}^*)|\boldsymbol{\theta}^*\bigr) \propto |\Sigma_\psi(\boldsymbol{\theta}^*)|^{-\frac{1}{2}}\exp\biggl(-\frac{1}{2}\bigl[\boldsymbol{\psi}(\boldsymbol{\theta}^*)-\boldsymbol{\mu}_\psi(\boldsymbol{\theta}^*)\bigr]^T\Sigma_\psi^{-1}(\boldsymbol{\theta}^*)\bigl[\boldsymbol{\psi}(\boldsymbol{\theta}^*)-\boldsymbol{\mu}_\psi(\boldsymbol{\theta}^*)\bigr]\biggr),$$

where $\boldsymbol{\mu}_\psi(\boldsymbol{\theta}^*)$ is a vector of conditional means, given by the Gaussian process emulators for $\mathbf{u}(\boldsymbol{\theta}^*)$, and $\Sigma_\psi(\boldsymbol{\theta}^*)$ 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, Richardson and Spiegelhalter (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 three 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

$$\begin{aligned}&\text{explained deviance}\\ &= 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{p}[I_y(\boldsymbol{\theta}_i,\mathbf{s}_j)\log(\hat{p}_{ij}) + (1-I_y(\boldsymbol{\theta}_i,\mathbf{s}_j))\log(1-\hat{p}_{ij})]}{\sum_{i=1}^{n}\sum_{j=1}^{p}[I_y(\boldsymbol{\theta}_i,\mathbf{s}_j)\log(\hat{p}_{0j}) + (1-I_y(\boldsymbol{\theta}_i,\mathbf{s}_j))\log(1-\hat{p}_{0j})]},\end{aligned} \tag{5.1}$$

where $\hat{p}_{ij} = \frac{\exp(\hat{\gamma}_{ij})}{1+\exp(\hat{\gamma}_{ij})}$ is an estimator for $P(I(\boldsymbol{\theta}_i,\mathbf{s}_j)=1)$ based on the fitted model with LPCA; $\hat{\gamma}_{ij}$ is an estimator for γ_{ij}, which is the (i,j)th element of $\boldsymbol{\Gamma}$ in (4.1); $\hat{p}_{0j} = \sum_{i=1}^{n} I(\boldsymbol{\theta}_i,\mathbf{s}_j)/n$ is the estimator for $P(I(\boldsymbol{\theta}_i,\mathbf{s}_j)=1)$, based on the "null model," which assumes a constant probability $P(I(\boldsymbol{\theta}_i,\mathbf{s}_j)=1) = p_{0j}$ 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:

$$\text{explained variance} = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{p}[Y(\boldsymbol{\theta}_i,\mathbf{s}_j) - \hat{Y}_{ij}]^2}{\sum_{i=1}^{n}\sum_{j=1}^{p}[Y(\boldsymbol{\theta}_i,\mathbf{s}_j) - \bar{Y}_j]^2}, \tag{5.2}$$

where $\hat{Y}_{ij}$ is the fitted value for $Y(\boldsymbol{\theta}_i,\mathbf{s}_j)$ 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 $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n$ at any $\mathbf{s}_j$ $(j=1,\ldots,p)$.

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 rectangular box in Figure S3 in which the input parameters are expected to have complicated effects, due to the complex terrain and interaction with the 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 nonleading components. For the principal components for the thickness, we set $J_w = 19$ to have a percentage of explained variation higher than 95% 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 five) 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 nonleading principal components. This suggests that, for this application, even the nonleading 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 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 231.05 m (while the overall mean ice thickness at those locations is 2117 m). 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 the calculation.

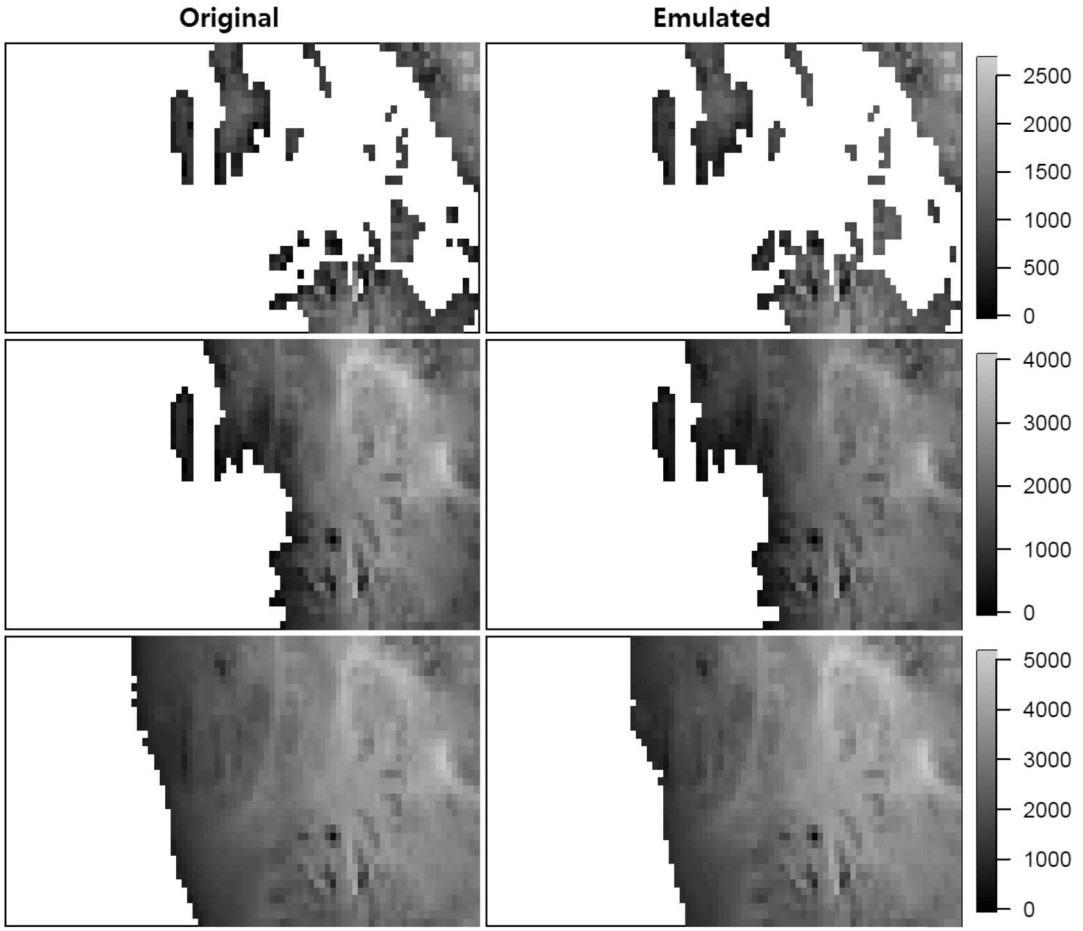

FIG. 2. *Examples of* 10*-fold cross-validation results, showing selected original spatial patterns from PSU-3D ice model* (*left column*) *and the corresponding emulated patterns* (*right column*). *The comparison shows that our emulator can predict the original model output with high accuracy.*

TABLE 1
10-*fold cross-validation results for two alternatives and our proposed emulator*

Method	# of LPCs	# of PCs	MAE
Partial Parallel	–	–	410.063
PCA only for Thickness	–	10	358.512
PCA only for Thickness	–	19	342.118
Proposed Emulator	13	19	231.050

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, Sacks and Welch (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 and Berger (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 a reduction in ice thickness below a certain threshold, an abrupt disintegration of the ice sheet in a certain part of the 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 $\boldsymbol{\theta}^*$. 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 nonnegativity without introducing a serious artifact due to transformation. While in the literature an exponential transformation is commonly used to enforce nonnegativity, 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 nonnegativity with only minimal distortion of data distribution:

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

This function preserves the original pattern of ice thickness by setting $h(\boldsymbol{\theta}_i, \mathbf{s}_j) = Y(\boldsymbol{\theta}_i, \mathbf{s}_j)$ for $Y(\boldsymbol{\theta}_i, \mathbf{s}_j) > 1$ m, while allowing the transformed variable to have negative values by setting $h(\boldsymbol{\theta}_i, \mathbf{s}_j) = \log(Y(\boldsymbol{\theta}_i, \mathbf{s}_j)) + 1$ for $0 < Y(\boldsymbol{\theta}_i, \mathbf{s}_j) \le 1$ 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 $q(\eta(\theta^*, \mathbf{s}) + \delta(\mathbf{s}) + \epsilon)$ 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:

(a) We first generate the assumed logit $\boldsymbol{\lambda}$, based on equation (4.6), by computing the predicted mean of $\boldsymbol{\xi}(\boldsymbol{\theta}^*)$ at the assumed truth $\boldsymbol{\theta}^*$ using the emulator constructed without the model run at $\boldsymbol{\theta}^*$. We then sample the coefficients for the discrepancy term $\mathbf{K}_v$, $\mathbf{v}$ from $N(\mathbf{0}, 0.4^2 I)$. 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 $\mathbf{s}_j$ we determine the ice presence and absence by sampling $I(\mathbf{s}_j)$ from a Bernoulli random variable with a success probability of $\frac{\lambda_j}{1+\lambda_j}$ where λ_j is the jth element of $\boldsymbol{\lambda}$.

(b) Second, we superimpose $\mathbf{K}_r\mathbf{r} + \boldsymbol{\epsilon}$ on $\mathbf{Y}(\boldsymbol{\theta}^*)$ at the assumed truth for $\boldsymbol{\theta}^*$, where $\mathbf{r} \sim N(0, 300^2 I)$ and $\boldsymbol{\epsilon} \sim N(0, 50^2)$. The variance 300^2 and 50^2 are chosen based on the modes of our prior distributions for σ_r^2 and σ_ϵ^2, defined in Section 4.2 *prior* part. After superimposing the synthetic errors, the locations with negative thickness values or $I(\mathbf{s}_j) = 0$ 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

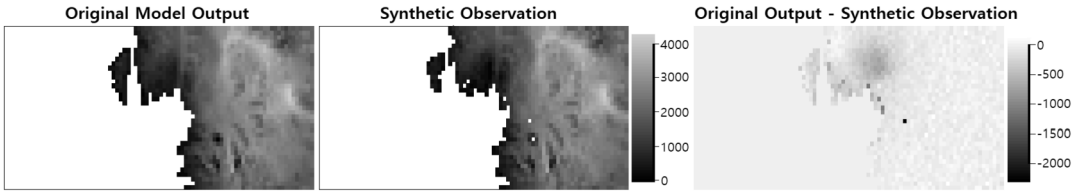

FIG. 3. *The map of the original model output for the assumed truth* (*left*), *the synthetic observation* (*middle*), *and the synthetic errors generated as described in Section* 5.4 (*right*). *The generated discrepancy terms remove some of the ice-covered regions and reduce the ice thickness in the high-discrepancy area*, *covered by the discrepancy matrices* $\mathbf{K}_r$ *and* $\mathbf{K}_v$ *described in Figure S*5.

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)).

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 of 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 the 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 toward the true value with the full thickness pattern. Since glaciologically LAPSE, CRHASE, LITH, 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

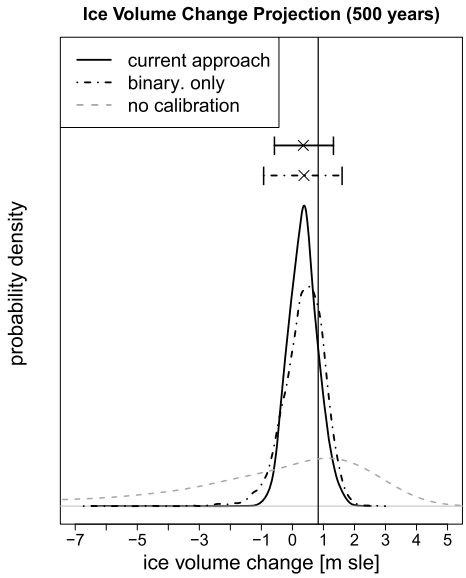

FIG. 4. *Ice volume change projections in sea level change equivalence* (*sle*) *based on the estimated posterior densities shown in Figures S*7 (*dashed and dotted line*), *S*8 (*solid line*), *and the original ensemble with* 499 *runs* (*i.e., no calibration applied, shown as dashed grey line*). *The 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*).

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, (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), and (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, that is, possible nonstationarity in the covariance structure for the GP emulator. We consider this direction as possible future work. The resulting projections are shown in Figure 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 $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ 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 Figure 4, is highly uncertain. Under this predictive distribution, both substantial decreases and substantial increases in WAIS volume are possible (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 toward lower values) was compensated by the positive bias in LAPSE (which will bias the prediction toward higher values) in the binary-only result, and hence improving on the modes of these parameters has

not led to a 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, that is, its bottom is below the sea level (see Figure 1(e)). 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 the 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 toward 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 have been shifted toward 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 toward 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 toward 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

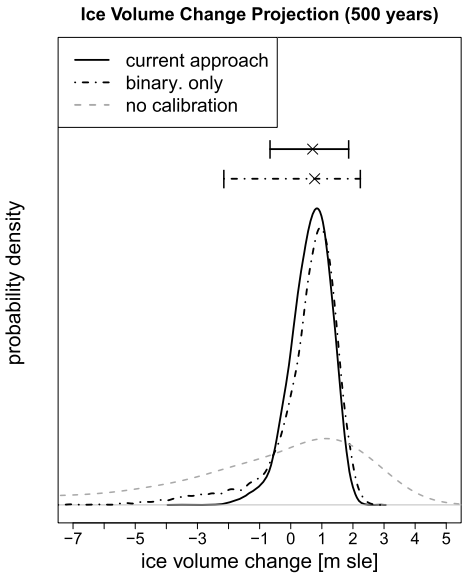

FIG. 5. *The same as Figure 4 except that the results are based on the densities in Figures S9 and S10, the posterior densities for observational data. Utilizing the full thickness pattern in calibration leads to reduced uncertainty in future projections compared to the result based only on the binary 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.*

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 Figure 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 $\mathbf{Z}^+$. To reduce the computational cost, we compute the test statistic for every 1000th 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 $m = 1558$ (the number of elements in $\mathbf{Z}^+$). 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 semicontinuous 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 semicontinuous 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 semicontinuous spatial data.

We use a mixture model for the semicontinuous 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 semicontinuous 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 semicontinuous data into binary data. In our real data application with the Bedmap2 dataset, we observe a 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 us 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, Bengio and Courville (2016)) that can extract low-dimensional features from a large number of predictors that are most relevant to the response variable.

APPENDIX: MATRIX COMPUTATION IN SECTION 4.3

Let $\mathbf{K}_+ = [\mathbf{K}_{+,u}\ \mathbf{K}_r]$; then the covariance matrix in (4.8) can be rewritten as

$$\begin{aligned}\Sigma_+ &= [\mathbf{K}_{+,u}\ \mathbf{K}_r]\Sigma_{\xi,r}[\mathbf{K}_{+,u}\ \mathbf{K}_r]^T + \sigma_\epsilon^2\mathbf{I}_m \\ &= \mathbf{K}_+\Sigma_{\xi,r}\mathbf{K}_+^T + \sigma_\epsilon^2\mathbf{I}_m.\end{aligned}$$

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

$$\left(\mathbf{K}_+\Sigma_{\xi,r}\mathbf{K}_+^T + \sigma_\epsilon^2\mathbf{I}_m\right)^{-1} = \sigma_\epsilon^{-2}\mathbf{I}_m - \sigma_\epsilon^{-2}\mathbf{K}_+\left(\Sigma_{\xi,r}^{-1} + \sigma_\epsilon^{-2}\mathbf{K}_+^T\mathbf{K}_+\right)^{-1}\mathbf{K}_+^T\sigma_\epsilon^{-2}.$$

This reduces the order of the computational cost of matrix inversion from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$. In a similar fashion and by applying the determinant formula (Harville (2008)), the determinant of the matrix can be rewritten as

$$\left|\mathbf{K}_+\Sigma_{\xi,r}\mathbf{K}_+^T + \sigma_\epsilon^2\mathbf{I}_m\right| = \sigma_\epsilon^{2m}\left|\Sigma_{\xi,r}^{-1} + \sigma_\epsilon^{-2}\mathbf{K}_+^T\mathbf{K}_+\right||\Sigma_{\xi,r}|.$$

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

Acknowledgments. The authors are grateful to Dr. David Pollard for the helpful discussion. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Funding. 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).

SUPPLEMENTARY MATERIAL

Supporting information (DOI: 10.1214/21-AOAS1577SUPP; .pdf). This is a PDF document containing additional discussions, figures and tables, cited as Chang et al. (2022).

REFERENCES

BAYARRI, M. J., BERGER, J. O., CAFEO, J., GARCIA-DONATO, G., LIU, F., PALOMO, J., PARTHASARATHY, R. J., PAULO, R., SACKS, J. et al. (2007). Computer model validation with functional output. *Ann. Statist.* **35** 1874–1906. MR2363956 https://doi.org/10.1214/009053607000000163

BERDAHL, M., LEGUY, G., LIPSCOMB, W. H. and URBAN, N. M. (2020). Statistical emulation of a perturbed basal melt ensemble of an ice sheet model to better quantify Antarctic sea level rise uncertainties. *Cryosphere* **15** 2683–2699.

BERGER, J. O., DE OLIVEIRA, V. and SANSÓ, B. (2001). Objective Bayesian analysis of spatially correlated data. *J. Amer. Statist. Assoc.* **96** 1361–1374. MR1946582 https://doi.org/10.1198/016214501753382282

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 https://doi.org/10.1002/env.2149

CAO, F., BA, S., BRENNEMAN, W. A. and JOSEPH, V. R. (2018). Model calibration with censored data. *Technometrics* **60** 255–262. MR3804253 https://doi.org/10.1080/00401706.2017.1345704

CHANG, W., HARAN, M., OLSON, R. and KELLER, K. (2014). Fast dimension-reduced climate model calibration and the effect of data aggregation. *Ann. Appl. Stat.* **8** 649–673. MR3262529 https://doi.org/10.1214/14-AOAS733

CHANG, W., HARAN, M., OLSON, R. and KELLER, K. (2015). A composite likelihood approach to computer model calibration with high-dimensional spatial data. *Statist. Sinica* **25** 243–259. MR3328813

CHANG, W., HARAN, M., APPLEGATE, P. and POLLARD, D. (2016a). Calibrating an ice sheet model using high-dimensional binary spatial data. *J. Amer. Statist. Assoc.* **111** 57–72. MR3494638 https://doi.org/10.1080/01621459.2015.1108199

CHANG, W., HARAN, M., APPLEGATE, P. and POLLARD, D. (2016b). Improving ice sheet model calibration using paleoclimate and modern data. *Ann. Appl. Stat.* **10** 2274–2302. MR3592057 https://doi.org/10.1214/16-AOAS979

CHANG, W., KONOMI, B. A., GEORGIOS, K., GUAN, Y. and HARAN, M. (2022). Supplement to "Ice model calibration using semicontinuous spatial data." https://doi.org/10.1214/21-AOAS1577SUPP

COOK, R. D. and NI, L. (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *J. Amer. Statist. Assoc.* **100** 410–428. MR2160547 https://doi.org/10.1198/016214504000001501

DE OLIVEIRA, V. (2005). Bayesian inference and prediction of Gaussian random fields based on censored data. *J. Comput. Graph. Statist.* **14** 95–115. MR2137892 https://doi.org/10.1198/106186005X27518

EDWARDS, T. L., BRANDON, M. A., DURAND, G., EDWARDS, N. R., GOLLEDGE, N. R., HOLDEN, P. B., NIAS, I. J., PAYNE, A. J., RITZ, C. and WERNECKE, A. (2019). Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature* **566** 58.

FRETWELL, P., PRITCHARD, H. D., VAUGHAN, D. G., BAMBER, J. L., BARRAND, N. E., BELL, R., BIANCHI, C., BINGHAM, R. G., BLANKENSHIP, D. D. et al. (2013). Bedmap2: Improved ice bed, surface and thickness datasets for Antarctica. *Cryosphere* **7** 375–393.

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740

GILKS, W. R., RICHARDSON, S. and SPIEGELHALTER, D. J., eds. (1995). *Markov Chain Monte Carlo in Practice. Interdisciplinary Statistics*. CRC Press, London. MR1397966 https://doi.org/10.1007/978-1-4899-4485-6

GLADSTONE, R. M., LEE, V., ROUGIER, J., PAYNE, A. J., HELLMER, H., LE BROCQ, A., SHEPHERD, A., EDWARDS, T. L., GREGORY, J. et al. (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.

GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773

GU, M. and BERGER, J. O. (2016). Parallel partial Gaussian process emulation for computer models with massive output. *Ann. Appl. Stat.* **10** 1317–1347. MR3553226 https://doi.org/10.1214/16-AOAS934

GU, M., PALOMO, J. and BERGER, J. O. (2019). RobustGaSP: Robust Gaussian stochastic process emulation in R. *R J.* **11** 112–136. https://doi.org/10.32614/RJ-2019-011

GU, M., WANG, X. and BERGER, J. O. (2018). Robust Gaussian stochastic process emulation. *Ann. Statist.* **46** 3038–3066. MR3851764 https://doi.org/10.1214/17-AOS1648

HARVILLE, D. A. (2008). *Matrix Algebra from a Statistician's Perspective*. Springer, Berlin.

HASTIE, T. J. (1992). Generalized additive models. In *Statistical Models in S* 249–307. Routledge, London.

HEATON, M. J., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D. et al. (2019). A case study competition among methods for analyzing large spatial data. *J. Agric. Biol. Environ. Stat.* **24** 398–425.

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 https://doi.org/10.1198/016214507000000888

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 https://doi.org/10.1111/1467-9868.00294

LE BROCQ, A. M., PAYNE, A. J. and VIELI, A. (2010). An improved Antarctic dataset for high resolution numerical ice sheet models (ALBMAP v1). *Earth Syst. Sci. Data* **2** 247–260.

LEE, S., HUANG, J. Z. and HU, J. (2010). Sparse logistic principal components analysis for binary data. *Ann. Appl. Stat.* **4** 1579–1601. MR2758342 https://doi.org/10.1214/10-AOAS327

LIU, Z., OTTO-BLIESNER, B., HE, F., BRADY, E., TOMAS, R., CLARK, P., CARLSON, A., LYNCH-STIEGLITZ, J., CURRY, W. et al. (2009). Transient simulation of last deglaciation with a new mechanism for Bølling–Allerød warming. *Science* **325** 310–314.

LOEPPKY, J. L., SACKS, J. and WELCH, W. J. (2009). Choosing the sample size of a computer experiment: A practical guide. *Technometrics* **51** 366–376. MR2756473 https://doi.org/10.1198/TECH.2009.08040

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. (2012). Description of a hybrid ice sheet-shelf model, and application to Antarctica. *Geosci. Model Dev.* **5** 1273–1295.

POLLARD, D., DECONTO, R. M. and ALLEY, R. B. (2015). Potential Antarctic Ice Sheet retreat driven by hydrofracturing and ice cliff failure. *Earth Planet. Sci. Lett.* **412** 112–121.

POLLARD, D., CHANG, W., HARAN, M., APPLEGATE, P. and DECONTO, R. (2016). Large-ensemble modeling of last deglacial and future ice-sheet retreat in the Amundsen Sea Embayment, West Antarctica. *Geosci. Model Dev.* **9** 1697–1723.

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

SALTER, J. M., WILLIAMSON, D. B., SCINOCCA, J. and KHARIN, V. (2019). Uncertainty quantification for computer models with spatial output using calibration-optimal bases. *J. Amer. Statist. Assoc.* **114** 1800–1814. MR4047301 https://doi.org/10.1080/01621459.2018.1514306

SANSÓ, B. and FOREST, C. (2009). Statistical calibration of climate system properties. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58** 485–503. MR2750089 https://doi.org/10.1111/j.1467-9876.2009.00669.x

STACKLIES, W., REDESTIG, H., SCHOLZ, M., WALTHER, D. and SELBIG, J. (2007). pcaMethods—A bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23** 1164–1167.

STEIN, M. L. (1992). Prediction and inference for truncated spatial data. *J. Comput. Graph. Statist.* **1** 91–110.

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.

SUNG, C.-L., HUNG, Y., RITTASE, W., ZHU, C. and WU, C. F. J. (2020). A generalized Gaussian process model for computer experiments with binary time series. *J. Amer. Statist. Assoc.* **115** 945–956. MR4107691 https://doi.org/10.1080/01621459.2019.1604361

TIPPING, M. E. and BISHOP, C. M. (1999). Probabilistic principal component analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 611–622. MR1707864 https://doi.org/10.1111/1467-9868.00196

WOODBURY, M. A. (1950). *Inverting Modified Matrices*. Princeton Univ., Princeton, NJ. Statistical Research Group, Memo. Rep. no. 42. MR0038136