The Pennsylvania State University
The Graduate School

COMPUTATIONAL METHODS FOR HIERARCHICAL SPATIAL MODELS AND ICE SHEET MODEL CALIBRATION

A Dissertation in
Statistics
by
Seiyon Lee

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2020

The dissertation of Seiyon Lee was reviewed and approved by the following:

Murali Haran
Professor of Statistics
Dissertation Advisor, Chair of Committee

Klaus Keller
Professor of Geosciences

Ben Shaby
Associate Professor of Statistics

Ephraim Hanks
Associate Professor of Statistics
Chair of Graduate Program

Abstract

Computer model calibration is a major component in projecting sea level rise and developing coastal flood-risk management strategies. Hierarchical spatial models have been used extensively to model spatially dependent observations across many fields such as climate science, ecology, public health, and epidemiology. The computational methods presented here have wide ranging applications in environmental sciences such as quantifying uncertainties in future sea level rise which are then used to formulate coastal risk management policies, and providing researchers from various fields with a fast and readily extendable approach to fit complex hierarchical spatial models of their choice. My dissertation research focuses on developing statistical and computational methods to address pressing issues in the environmental sciences. My contributions are as follows: (1) a fast particle-based approach for calibrating a three-dimensional Antarctic ice sheet model. I developed a sequential Monte Carlo method that leverages the massive parallelization inherent to modern high-performance computing systems; (2) an efficient and extendable approach for fitting high-dimensional hierarchical spatial models. I propose a discretized and dimension-reduced representation of the underlying spatial random field using empirical basis functions on a triangular mesh; and (3) a computationally efficient method for modeling high-dimensional zero-inflated spatial observations.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisors, Murali Haran and Klaus Keller, for all their help and guidance. I am grateful to have been mentored by these two outstanding researchers and even better human beings. I thank Murali for his endless patience, dedication to his craft, and integrity as a scholar. Thank you for always making time for your students and providing a sense of calm amidst the chaos of research. I am grateful to Klaus for providing fruitful collaborations, supporting me through research assistantships, and consistently providing chocolates at our weekly meetings. Thank you for sharing your high-level/big-picture vision and showing me how research is "more than just the methods."

I would also like to thank Ephraim Hanks and Ben Shaby for serving as my dissertation committee members and helpful discussion with my research. A special thanks to David Pollard and Rob Fuller for their tremendous help in the Antarctic ice sheet calibration research. I am grateful to the Keller research group, particularly Vivek Srikrishnan, Tony Wong, Casey Hegelson, Joel Roop-Eckart, and Sanjib Sharma, for the stimulating weekly discussions. I would also like to thank my fellow students and friends, who made my time at Penn State quite memorable, including Jaewoo Park, Yawen Guan, Xiaoxiao Li, Kyongwon Kim, Likun Zhang, Christian Schmid, and Nick Sterge. Thank you to Haim Lee for all the moral support and encouragement, especially in the homestretch.

Last but not least, I must thank my parents and anchors, Dr. Munhee Lee and Myongja Emily Lee, for their unwavering support and love. Thank you for setting me on this path. None of this would have been possible without you.

This dissertation was partially supported by the by the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 and the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Earth and Environmental Systems Modeling, MultiSector Dynamics, Contract No. DE-SC0016162. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect

the views of the US Department of Energy or the National Science Foundation. I would also like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

Chapter 1

Introduction

Complex computer models play an increasingly important role in the environmental sciences. For example, state-of-the-art Antarctic ice sheet computer models (Pollard and DeConto, 2012a), or simulators, help us understand the ice dynamics and long-term behavior of ice sheets. Antarctic ice sheet models play an prominent role in projecting sea level rise, which is critically tied to the sustainability of highly populated metropolitan areas, resource allocation and management, and risks to life and property due to flood hazards. In the environmental sciences, recent data collection initiatives have lead to increasingly sophisticated and high-dimensional spatial datasets. For example, state government agencies conduct large-scale field surveys (Maryland's Mapping and GIS Data Portal, 2018) at carefully sampled locations to monitor watershed resources. In addition, satellites traverse the globe measuring air pollution in the form of remotely-sensed aerosol optical depth (AOD) and collect massive amounts of spatially-indexed data at high-resolutions (Murray et al., 2019).

These recent advances in scientific modeling and data collection uncovered novel challenges pertaining to data storage, computationally intensive computer model simulations, projections with deep uncertainties, and increasingly complex spatial models. These challenges provide exciting research opportunities that lie in the intersection of statistics, environmental sciences, and computation. My dissertation focuses on developing computationally efficient statistical methods to address two key scientific problems: (1) incorporating information from various data sets to tune, or "calibrate," complex computer models to enable a better understanding

of the past, present, and future of the climate system; and (2) developing fast algorithms to model complex spatial datasets from the environmental sciences. The methods developed here have potentially far reaching applications across many disciplines.

My dissertation research consists of three projects:

1. **Fast Particle-based Approach for Computer Model Calibration:** Complex ice sheet computer models play a prominent role in climate science, particularly in projecting future sea level rise and informing coastal flood-risk management strategies. These models require parameters that are calibrated, or tuned, based on observations and prior knowledge. For many computer models, existing calibration methods are either computationally prohibitive or largely underestimate parametric uncertainty. I propose a sequential Monte Carlo-based calibration method that provides good approximations with shorter calibration wall times. This enables important computer model experiments that have been computationally infeasible using current calibration approaches.

2. **Projection-based Intrinsic Conditional Autoregression (PICAR):** High-dimensional hierarchical spatial models are widely used across many disciplines, for instance, species abundance in ecology, ice presence in glaciology, geo-referenced survey responses in public health studies, and crime incidence in urban areas. Examples of these models include spatial generalized linear mixed models (SGLMMs), spatially varying coefficient models, spatial ordinal response models, and two-part models for zero-inflated spatial data. The high-dimensional observations pose computational challenges for model fitting such as costly matrix operations and slow mixing Markov Chain Monte Carlo algorithms (MCMC). I propose a projection-based intrinsic conditional autoregression (PICAR) approach to reduce the dimensions and also de-correlate the spatial random effects.

3. **High-dimensional Zero-inflated Spatial Models:** Zero-inflated spatial observations are spatially dependent data containing a large proportion of zeros. These types of data are prevalent across many disciplines such as climate science, ecology, infectious disease modeling, criminology, and health

and human services. Traditional spatial models may not be suitable for modeling zero-inflated spatial data due to the excess zeros and high levels of over/underdispersion. Moreover, fitting high-dimensional spatial models can be computationally prohibitive. Here, I propose a computationally efficient method for fitting high-dimensional zero-inflated spatial models. In addition, I provide practical guidelines for selecting the appropriate class of zero-inflated models as well as demonstrate this approach of simulated and real-world examples.

In the remainder of this chapter, I provide an overview of the main areas of focus along with a brief literature review.

1.1 Fast Computer Model Calibration

In this section, I introduce the scientific motivation and general framework behind computer model calibration. I also provide a brief review of existing methods and propose the basis for my fast particle-based approach for calibrating complex computer models.

1.1.1 Scientific Motivation

Substantial mass loss of the Antarctic ice sheet poses considerable challenges to coastal-flood risk management. Sound coastal flood-risk management strategies rely on sea level projections as well as quantifying their associated uncertainties. The Antarctic ice sheet remains the single largest source of uncertainty for future sea level rise (DeConto and Pollard, 2016). Geological records suggest that ice sheets can substantially drive global sea level rise (Deschamps et al., 2012) possibly as high as 58 m (Fretwell et al., 2012). Nearly eight percent of the current global population is threatened by a five meter rise in sea level (Nicholls et al., 2008) and 13 percent of the global urban population is threatened by a rise of ten meters (McGranahan et al., 2007). However, sea level rise projections are based on deeply uncertain projections of the Antarctic ice sheet's mass loss (Le Bars et al., 2017; Wong et al., 2017; Le Cozannet et al., 2017). Therefore, quantifying and characterizing the long-term behavior of the Antarctic ice sheet is central to

designing coastal risk management strategies (cf. Garner and Keller, 2018; Sriver et al., 2018; Oppenheimer and Alley, 2016).

Ice sheet computer models are deterministic mathematical models that simulate dynamical ice processes such as the long-term mass loss. However, ice sheet models rely on poorly constrained parameters, and recent studies show that this parametric uncertainty results in highly uncertain projections of sea level change (Stone et al., 2010; Applegate et al., 2012; Fitzgerald et al., 2012; Collins, 2007), which inevitably affects climate risk decision-making (O'Neill et al., 2006; Hannart et al., 2013). Recent studies have addressed this parametric uncertainty via calibration studies using modern observations, but these are either limited to simple ice sheet models (Ruckert et al., 2017; Fuller et al., 2017) or a small number of model parameters (Chang et al., 2016b; Edwards et al., 2019; Schlegel et al., 2018). In Chapter 2, I propose a particle-based approach to calibrate the Pennsylvania State University three-dimensional Antarctic ice sheet model (PSU3D-ICE) (Pollard and DeConto, 2012a); thereby characterizing and quantifying key deep uncertainties surrounding sea level projections.

1.1.2 Overview of Calibration

Computer models are deterministic mathematical models that output simulations of real-world physical processes. The underlying mathematical models are comprised of complex systems of differential equations, which are constructed based on scientific understandings of the physical processes. Direct experimentation using real world processes may be impractical (e.g. climate processes), so computer models provide a more viable and cheaper alternative to generate realizations. Computer models have been used to model dynamic processes across many disciplines including weather forecasting, glaciology, ecology, epidemiology, industrial engineering, sociology, and economics.

Two major components of computer models are the model outputs and the model input parameters. The model outputs may be a scalar value, time series, or a spatial field, particularly in the environmental sciences. The model inputs are a collection of parameters required to run the computer model. The computer models considered here are deterministic, meaning that running the computer model

with same input parameters will always generate the same output. In this dissertation, the computer models are treated as "black-box" models where the internal mathematical models are not manipulated.

In order to accurately represent the physical processes, the model input parameters must be calibrated, or "tuned," by comparing the generated model outputs to the observations. In the past, calibration consisted of 'plug-in' approaches to select the input parameter set whose corresponding model outputs best fit the observed data. However, 'plug-in' methods may inaccurately represent the system of interest as they ignore various sources of uncertainties (e.g., parametric, observational errors, and systematic model-observation discrepancies) (Kennedy and O'Hagan, 2001). In this section, I introduce the Bayesian calibration framework (Kennedy and O'Hagan, 2001) designed to characterize and quantify the numerous sources of uncertainty and provide a brief review of existing calibration methods.

1.1.2.1 General Framework

In computer model calibration, key computer model parameters are inferred by comparing the computer model output and observational data (cf. Kennedy and O'Hagan, 2001; Bayarri et al., 2007). Moreover, calibration also accounts for important sources of uncertainty such as the model-observation discrepancy and observational error (Kennedy and O'Hagan, 2001; Bayarri et al., 2007; Brynjarsdottir and O'Hagan, 2014). Model-observation discrepancy is the systematic difference between the observations and model outputs attributed to the computer model's misrepresentation of the physical processes. The observational errors represent the non-systematic measurement errors.

Kennedy and O'Hagan (2001) presents the general Bayesian framework for computer model calibration. The unknown input parameters θ are represented as random variables with prior distribution $p(\theta)$, and the posterior distribution $\pi(\theta|Z)$ is obtained by assimilating the observed data Z. Ultimately, $\pi(\theta|Z)$ characterizes the parametric uncertainty in calibration problems.

Here, I introduce the Bayesian calibration framework for a generic computer model whose output is a spatial random field. Let $Y(s, \theta)$ be the computer model output at the spatial location $s \in \mathcal{S} \subseteq \mathbb{R}^2$ and the parameter setting $\theta \in \Theta \subseteq \mathbb{R}^d$. $\mathcal{S}$ is the spatial domain of the process, and Θ is the parameter space of

the computer model with integer d being the number of input parameters. $\mathbf{Y} = (Y(s_1, \theta_i), ..., Y(s_n, \theta_i))^T$ is the computer model output at parameter setting θ_i and spatial locations $(s_1, ..., s_n)$. $\mathbf{Z} = (Z(s_1), ..., Z(s_n))^T$ is the observed spatial process at locations $(s_1, ..., s_n)$.

The observational data Z is modeled as:

$$Z = Y(\theta) + \delta + \epsilon, \tag{1.1}$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$ are the independently and identically distributed observational errors, and δ is the systemic data-model discrepancy term. The discrepancy term δ is generally modeled as a zero-mean Gaussian process, where $\delta \sim N(0, \Sigma_\delta(\xi_\delta))$. $\Sigma_\delta(\xi_\delta)$ is the spatial covariance matrix between spatial points $s_1, ..., s_n$ with covariance parameters ξ_δ. Prior distributions are chosen for the model parameters, θ, and observational error variance, σ_ϵ^2. On the other hand, informative priors are necessary for the discrepancy term's covariance parameters ξ_δ. Then, θ, σ^2, and ξ_δ are inferred by sampling from the posterior distribution, $\pi(\theta, \sigma_\epsilon^2, \xi_\delta | Z)$, via Markov Chain Monte Carlo (MCMC).

The hierarchical framework for computer model calibration is as follows:

$$\begin{aligned}
&\textbf{Data Model:} && \mathbf{Z}|\boldsymbol{\theta}, \delta \sim \mathcal{N}(Y(\boldsymbol{\theta}) + \delta, \sigma_\epsilon^2 I), \\
&\textbf{Process Model:} && \delta|\xi_\delta \sim \mathcal{N}(0, \Sigma_\delta(\xi_\delta)), \\
&\textbf{Parameter Model:} && \boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \quad \xi_\delta \sim p(\xi_\delta), \quad \sigma_\epsilon^2 \sim p(\sigma_\epsilon^2)
\end{aligned} \tag{1.2}$$

The observation-model discrepancy term δ represents the systematic differences between the model outputs and observations. The discrepancy term characterizes model inadequacy, or the systematic difference between the mode output and observations. To illustrate, even if the computer model were run at the best possible input parameter settings, there would inevitably be a difference between the computer model output and the actual observations. This can be attributed to incomplete specifications of processes or model inadequacy (Higdon et al., 2008). This discrepancy term is essential for parameter calibration (Bhat et al., 2010; Bayarri et al., 2007) and ignoring it may yield biased and overconfident estimates and projections (Brynjarsdottir and O'Hagan, 2014).

1.1.3 Existing Calibration Methods

The Bayesian calibration approach can be computationally prohibitive for complex computer models. Depending on the complexity of the physical processes, a single evaluation of the computer model may incur high computational costs (e.g., long model run times and use of multiple processors). Under the Bayesian calibration framework, computer models must be run repeatedly at various parameter (input) settings to accurately assess the underlying uncertainties (parameter and discrepancy). Often, the computer model is too expensive to be embedded in the resulting MCMC algorithm as it requires repeated computer model evaluations.

1.1.3.1 Emulation-Calibration

To counteract this challenge, emulation-calibration calibration approaches (Sacks et al., 1989) have been developed where surrogate models or 'emulators' replace the more expensive expensive computer model. The surrogate models are a computationally efficient approximation of the expensive original computer model. The two most common surrogate modeling approaches are Gaussian process emulation (Sacks et al., 1989; Currin et al., 1991) and polynomial chaos expansions(Ghanem and Spanos, 1991; Xiu and Karniadakis, 2002). Other surrogate modeling techniques include machine learning methods using support vector machines (Ciccazzo et al., 2014; Pruett and Hester, 2016) and neural networks (Eason and Cremaschi, 2014; Gorissen et al., 2009).

For emulation-calibration approaches, the computer model output $Y(\theta)$ is replaced with the surrogate model output $\eta(\theta)$ for input parameter θ. Calibration proceeds similarly to the original case. Here, the observational data Z is modeled as follows,

$$Z = \eta(\theta) + \delta + \epsilon, \tag{1.3}$$

where ϵ and δ are the observational errors and model discrepancy term, respectively. The Bayesian hierarchical framework is similar to Equation 1.2; however, the computer model output $Y(\theta)$ is replaced by the emulator output $\eta(\theta)$.

$$
\begin{aligned}
&\textbf{Data Model:} && \mathbf{Z}|\boldsymbol{\theta}, \delta \sim \mathcal{N}(\eta(\boldsymbol{\theta}) + \delta, \sigma_\epsilon^2 I), \\
&\textbf{Process Model:} && \delta|\xi_\delta \sim \mathcal{N}(0, \Sigma_\delta(\xi_\delta)), \\
&\textbf{Parameter Model:} && \boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \quad \xi_\delta \sim p(\xi_\delta), \quad \sigma_\epsilon^2 \sim p(\sigma_\epsilon^2)
\end{aligned} \tag{1.4}
$$

Gaussian Process Emulators

Gaussian process emulation (Sacks et al., 1989; Currin et al., 1991) is a surrogate modeling approach popular within the statistics community. Here, the surrogate model output, or emulator, is treated as a realization from a stochastic process. Let $\mathbf{Y} = \{Y(\theta_1), ..., Y(\theta_p)\}$ be a collection of model runs evaluated at p design points $\theta_1, ..., \theta_p$. Guassian process emulators interpolate the model outputs via a Gaussian process in the parameter space Θ. The emulator is constructed as:

$$\mathbf{Y} \sim \mathcal{N}(X\beta, \Sigma(\xi)),$$

where $\mathbf{X}$ is a $p \times b$ matrix of covariates and $\Sigma(\xi)$ is the $p \times p$ covariance matrix. The covariate matrix $\mathbf{X}$ may include functions of the input parameters θ or set to be 0. The covariance matrix is defined as $\Sigma(\xi)_{ij} = C(\theta_i, \theta_j; \xi)$, where $C(\cdot)$ is a covariance function (e.g., the Matérn class). The Gaussian process emulator is constructed by estimating parameters β and ξ. This framework assumes that the model output at untried settings $Y(\theta^*)$ is normally distributed when conditioned on $\mathbf{Y}$. The emulator interpolates the model output $Y(\theta^*)$ at unknown settings θ^* such that the emulator output $\eta(\theta^*) = E[Y(\theta)^*|\mathbf{Y}]$ or the predictive process of Y. To illustrate, suppose the joint distribution of $Y(\theta^*)$ and $\mathbf{Y}$ is:

$$\begin{bmatrix} Y(\theta^*) \\ \mathbf{Y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_{11}(\xi) & \Sigma_{12}(\xi) \\ \Sigma_{21}(\xi) & \Sigma_{22}(\xi) \end{bmatrix} \right),$$

, where $\Sigma_{22}(\xi)$ is the covariance matrix of the simulator output run at the initial settings, $\Sigma_{11}(\xi)$ is the covariance matrix of the simulator output at the untried setting θ^*, and $\Sigma_{12}(\xi)$ and $\Sigma_{21}(\xi)$ are the cross-covariance between the computer model runs at the untried settings and the untried settings. Then, the emulator output at the untried input settings $\eta(\theta^*)$ is the conditional mean of $Y(\theta)^*$ given

$\mathbf{Y}$:

$$\eta(\theta^*) = \Sigma_{12}(\xi)\Sigma_{22}^{-1}(\xi)\mathbf{Y}$$

Polynomial Chaos Expansions

Polynomial chaos expansions (Ghanem and Spanos, 1991; Xiu and Karniadakis, 2002) is another widely used surrogate modeling approach stemming from the applied mathematics community. Here, the surrogate model output is generated using a series expansion of orthogonal polynomial basis functions. Similar to the Gaussian process emulation approach, there exists a training set of model outputs $\mathbf{Y} = \{Y(\theta_1), ..., Y(\theta_p)\}$ evaluated at p design points $\theta_1, ..., \theta_p$. Suppose the vector of model parameters $\boldsymbol{\theta} \in \mathbb{R}^d$ has a given probability density function $f_{\boldsymbol{\theta}}(\theta) = \prod_{i=1}^{d} f_{\theta_i}(\theta_i)$, where $f_{\theta_i}(\theta_i)$ denotes the marginal probability density function of input θ_i with support Θ_i. For the calibration framework in Equation 1.2, $f_{\theta_i}(\theta_i)$ is the prior distribution $p(\theta_i)$.

Assuming that the computer model output $Y(\theta)$ has finite variance $\text{Var}[Y(\theta)] < \infty$, $Y(\theta)$ can be approximated as a series expansion of polynomial basis functions (polynomial chaos expansion):

$$Y(\theta) \approx \sum_{i=1}^{m} a_i \psi_i(\theta),$$

where $\psi_i(\theta), \quad i = 1, ..., m$ are the orthogonal polynomial basis functions and $\mathbf{a} = \{\mathbf{a}_1, ..., \mathbf{a}_m\}$ are the basis coefficients. The orthogonal polynomial basis functions $\boldsymbol{\Psi} = \{\psi_1(\theta), ..., \psi_m(\theta)\}$ are constructed based on the probability distribution of the model input parameters $f_{\boldsymbol{\theta}}(\theta)$ (Table 1.1.3.1). For example, a set of model input parameters θ whose marginal distributions (prior distributions) are uniform (e.g., $f_\theta(\theta) \sim \text{Unif}(a, b)$ and $\theta \in (a, b)$) would correspond to orthogonal polynomial basis functions $\boldsymbol{\Psi}$ from the Lengendre polynomial family.

Polynomial weights $\mathbf{a}$ can be estimated by least-squares regression (Isukapalli et al., 1998; Berveiller et al., 2006), where the response variables are the model outputs $Y(\theta)$ and the covariates are the polynomial basis functions $\psi(\theta)$ from the training set. The polynomial weights $\mathbf{a}$ can be represented as the numerical solution of an integral (Eldred et al., 2008), which typically does not have an

Table 1.1. Orthogonal polynomial families based on dsitribution of model input parameters. Adapted from Owen (2017).

Input/Parameter Type	Input/Parameter Distribution $f_{\boldsymbol{\theta}}(\theta)$	Polynomial Family	Support
Continuous	Gaussian	Hermite	$(-\infty, \infty)$
	Gamma	Laguerre	$(0, \infty)$
	Beta	Jacobi	(a, b)
	Uniform	Legendre	(a, b)
Discrete	Poisson	Charlier	$\{0, 1, ...\}$
	Binomial	Krawtchouk	$\{0, 1, ..., N\}$
	Negative binomial	Meixner	$\{0, 1, ..., N\}$
	Hypergeometric	Hahn	$\{0, 1, ..., N\}$

analytical solution. Hence, Monte Carlo sampling methods (Ghiocel and Ghanem, 2002; Reagana et al., 2003) and quadrature methods (Le Maitre et al., 2002; Eldred et al., 2008) have been used for numerical integration. The surrogate model output at the untried input settings $\eta(\theta^*)$ is a linear combination of the polynomial basis functions using the estimated basis coefficients $\hat{\mathbf{a}}$:

$$\eta(\theta^*) = \sum_{i=1}^{m} \hat{\mathbf{a}}_i \psi_i(\theta^*)$$

Limitations

Though Gaussian process emulators and polynomial chaos expansions are widely used surrogate models, both methods have their strengths and weaknesses. In cases with large training samples, fitting the Gaussian process emulator can be computationally expensive (Owen, 2017; O'Hagan, 2013) as it involves repeatedly evaluating a high-dimensional multivariate normal likelihood function. Constructing polynomial chaos expansions tend to be computationally efficient since: (1) a finite collection of polynomial basis functions are selected a priori; (2) basis coefficients are readily estimated using regression or spectral projection; and (3) the surrogate model output (interpolation) is represented as a linear combination of the basis functions and coefficients. On the other hand, Gaussian processes emulators tend to be more flexible than polynomial chaos expansions and may be better suited for computer models displaying non-linear behavior (Owen, 2017). While polynomial chaos expansions only provide the surrogate model output, Gaussian

process emulators generate both the model output as well as the approximation uncertainty of the output. Moreover, Gaussian process emulators can quantify model fidelity (Owen, 2017; O'Hagan, 2013) by using the information in the covariance function.

Both surrogate modeling approaches suffer from model infidelity where the surrogate model output $\eta(\theta)$ may poorly approximate the true computer model output $Y(\theta)$. In general, these surrogate models are trained using a limited number of computer model runs (Sacks et al., 1989). Dense sampling schemes, such as full factorial or fractional factorial designs, may help capture higher order interactions; however, running the computer model at each of the design points is costly. Space-filling designs such as the Latin Hypercube Design (McKay et al., 2000; Steinberg and Lin, 2006; Stein, 1987) or adaptive experimental designs (Chang et al., 2016a; Gramacy and Apley, 2015; Urban and Fricker, 2010; Queipo et al., 2005) use fewer design points, but may possibly generate low-fidelity surrogate models by ignoring higher order interactions among inputs (Liu and Guillas, 2017). Some studies have employed emulation-calibration methods (Sansó et al., 2008; Liu et al., 2009a; Bhat et al., 2010) to calibrate computer models with long run times, but these approaches are applicable to only a small number of parameters. For computer models with longer run times and a large number of model parameters, emulation-calibration can be computationally prohibitive because building an accurate emulator requires a large training data set (Bastos and O'Hagan, 2009; Maniyar et al., 2007).

1.1.3.2 Particle-based Approaches

The limitations of existing calibration methods motivate the development of new calibration approaches for complex computer models with a moderate model run times (6 seconds to 15 minutes) and a moderate number of input parameters (5-20). The long single model run times prohibit calibration under the Kennedy and O'Hagan (2001) framework and the large number of model parameters presents challenges for emulation-calibration methods. Examples of such computer models include a coarser resolution Antarctic ice sheet model (Pollard and DeConto, 2012a), single column atmospheric models (Bony and Emanuel, 2001; Dal Gesso and Neggers, 2018; Gettelman et al., 2019), hydrological soil moisture models

(Sorooshian et al., 1993; Liang et al., 1994), simplified earth systems models (Monier et al., 2013), and integrated multi-Sector models for human and earth dynamics (Kim et al., 2006).

Sequential Monte Carlo (SMC), or particle-based, approaches (Del Moral et al., 2006; Doucet et al., 2000; Liu and West, 2001; Chopin, 2002; Crisan and Doucet, 2000) have gained wide practical use in uncertainty quantification (cf. Kantas et al., 2015; Papaioannou et al., 2016; Kalyanaraman et al., 2016; Jeremiah et al., 2011; Morzfeld et al., 2018; Higdon et al., 2008). In the model calibration context, SMC approaches approximate the posterior distribution $\pi(\theta|Z)$ (Equation 1.2) using a weighed set of samples from a different distribution that may be easier to draw sample from. Since much of the SMC operations are embarrassingly parallel, these methods are well-suited for modern high-performance computing systems. By leveraging massive parallelization across multiple processors, particle-based approaches can drastically reduce calibration walltimes (Kalyanaraman et al., 2016).

Sequential Monte Carlo algorithms use sampling-importance-resampling (Gordon et al., 1993; Doucet et al., 2001), a popular method to approximate a target distribution $\pi(\theta)$ using particles, or samples, from an importance distribution $q(\theta)$. Sampling-importance-resampling generates an empirical distribution using weighted samples from $q(\theta)$, where the weights are calculated using importance sampling. Importance sampling is a general technique used to estimate $\mu = E_\pi\big[g(\theta)\big]$ where $g(\theta)$ is a function of θ. Given $q(\theta) > 0$ whenever $g(\theta)\pi(\theta) > 0, \quad \forall\theta \in \Theta$, it follows that $E_\pi\big[g(\theta)\big] = E_q\Big[g(\theta)w(\theta)\Big]$, where $w(\theta) = \frac{\pi(\theta)}{q(\theta)}$ is the importance weight and $\sum_{i=1}^{N} w(\theta_i) = 1$. Here, $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{N} g(\theta_i)w(\theta_i)$ is the importance sampling estimator and $\hat{\mu}_n \to \mu$ with probability 1 by the strong law of large numbers. Next, sampling-importance-sampling approximates the target distribution $\pi(\theta)$ using the empirical distribution of the samples $\hat{\pi}(\theta)$, and their corresponding normalized weights $\tilde{w}(\theta_i)$'s:

$$\pi(\theta) \approx \hat{\pi}(\theta) = \sum_{i=1}^{N} \tilde{w}(\theta_i)\delta(\theta_i),$$

where $\delta(\theta_i)$ is the Dirac measure that puts unit mass at θ_i and $\sum_{i=1}^{N} \tilde{w}(\theta_i) = 1$.

Poor choices of $q(\theta)$ may yield inaccurate approximations of the target distribution (Doucet et al., 2000) due to weight degeneracy and sample impoverishment.

Mixture approximations (Gordon et al., 1993) or kernel smoothing methods (Liu and West, 2001) have been used to mutate or rejuvenate the replicated particles. However, these methods may not scale well to high-dimensional target distributions (Doucet et al., 2000). Adaptive tempering schedules and mutation stages (Jasra et al., 2011) using the Metropolis-Hastings transition kernel (Gilks and Berzuini, 2001), genetic algorithms (Zhu et al., 2018) or different transition kernels, $K(\cdot)$ (Papaioannou et al., 2016; Murray et al., 2016) have been implemented to address these challenges.

In Chapter 2, I propose a fast particle-based calibration approach designed for computer models with moderate model run times (6 seconds to 15 minutes) and moderate number of input parameters (5-20), namely the Pennsylvania State University three-dimensional Antarctic Ice Sheet Model. This approach reduces calibration wall times by (1) parallelizing the bulk of computer model runs; (2) imposing stopping rules within the algorithm; and (3) applying adaptive sampling techniques to limit expensive model evaluations. Within the context of sea level projections, this method better characterizes parametric and projection uncertainty than existing approaches. In addition, the particle-based approach enables important computer model experiments that were once computationally prohibitive.

1.2 Hierarchical Spatial Models

Advances in computing and spatial data collection have enabled researchers to construct increasingly complex spatial models to represent various environmental processes. These models may use multiple data sources, account for various sources of uncertainties or errors, and include more than one layer of latent, or hidden, spatial processes. Bayesian hierarchical spatial models are a popular class of spatial models that provide a flexible framework designed to account for these complex features. This class of models are commonly used to model complex spatial observations across many fields; for example, species abundance in ecology, ice presence in glaciology, geo-referenced survey responses in public health studies, and crime incidence in urban areas.

Bayesian hierarchical spatial models are characterized by a hierarchy of condi-

tional distributions often broken up into three component models (Berliner, 1996; Gelfand et al., 2003) - data, process, and parameter models. The overall structure is as follows:

Data Model:	Data | Process, Parameters	
Process Model:	Process | Parameters	(1.5)
Parameter Model:	Parameters and Hyper-parameters	

In the first stage, the data model is the probability distribution of the observations conditioned on the underlying spatial processes and model parameters. Here, the data likelihood function usually serves as the data model. In the second stage, the process model represents the underlying, often latent, spatial processes. The process model typically consists of high-dimensional multivariate probability distributions and in some cases, nested layers of sub-processes. Finally, the third stage includes the prior distributions of the model parameters and their corresponding hyper-parameters.

1.2.1 Hierarchical Modeling Framework

Let $Z(s)$ denote the observed data at location s in a spatial domain $\mathcal{D} \subset \mathbb{R}^d$ where d is generally 2 or 3. $Z(s)$ is defined as:

$$Z(s) = X(s)\beta + w(s) + \epsilon(s), \text{ for s} \in D, \tag{1.6}$$

where $X(s)$ is a set of k covariates associated with location s and β is a k-dimensional vector of coefficients. The micro-scale measurement errors or nugget are modeled as an uncorrelated Gaussian process with zero mean and variance τ^2 where $\epsilon(s) \sim N(0, \tau^2)$ for all $s \in \mathcal{D}$.

Spatial dependence is introduced by modeling the spatial random effects $\mathbf{W} = \{w(s) : s \in \mathcal{D}\}$ as a stationary zero-mean Gaussian process with a positive definite covariance function $C(\cdot)$. For a finite set of locations $s = (s_1, ..., s_n)$, the spatial random effects $\mathbf{W}$ are distributed as a multivariate normal distribution $\mathbf{W}|\Theta \sim N(0, C(\Theta))$ with covariance function parameters Θ and the covariance matrix $C(\Theta)$ where $C(\Theta)_{ij} = \text{Cov}(w(s_i), w(s_j))$. The Matérn covariance function

is a widely used class of stationary and isotropic covariance functions (Stein, 2012) with parameters $\Theta = (\sigma^2, \phi, \nu)$ such that:

$$C(s_i, s_j) = \sigma^2 \frac{1}{\Gamma(\nu)2^{\nu-1}} \Big(\sqrt{(}2\nu)\frac{h}{\phi}\Big)^{\nu} K_\nu\Big(\sqrt{(}2\nu)\frac{h}{\phi}\Big),$$

where $R(\phi)$ is the correlation matrix, $h = ||s_i - s_j||$ is the Euclidean distance between locations s_i and s_j, $\sigma^2 > 0$ is the partial sill or scale parameter of the process, and $\phi > 0$ is the range parameter for spatial dependence. $K_\nu(\cdot)$ is the modified Bessel function of the second kind where the smoothness parameter ν is commonly fixed prior to model fitting.

For Gaussian observations, hierarchical spatial models may be broadly described as (cf. Wikle et al., 1998):

$$\begin{aligned}
&\textbf{Data Model:} && Z(s)|\beta, \mathbf{W}, \tau^2 \sim N(X(s)\beta + w(s), \tau^2) \\
&\textbf{Process Model:} && \mathbf{W}|\phi, \sigma^2 \sim N(0, \sigma^2 R_\phi), \qquad \mathbf{W} = \{w(s_1), ..., w(s_n)\} \\
&\textbf{Parameter Model:} && \beta \sim p(\beta), \phi \sim p(\phi), \sigma^2 \sim p(\sigma^2), \tau^2 \sim p(\tau^2)
\end{aligned} \tag{1.7}$$

Generalized Linear Spatial Models

Non-Gaussian spatial observations are typically modeled using spatial generalized linear mixed models (SGLMMs) (Diggle et al., 1998, Haran, 2011). Let $\{Z(s) : s \in \mathcal{D}\}$ be a non-Gaussian spatial random field. Assuming $Z(s)$ are conditionally independent given the latent random spatial field $\mathbf{W}$, the conditional mean $E[Z(s)|\beta, \mathbf{W}, \epsilon(s)]$ can be modeled through a linear predictor $\eta(s)$:

$$\eta(s) = g\{E[Z(s)|\beta, \mathbf{W}], \epsilon(s)\} = X(s)\beta + w(s) + \epsilon(s),$$

where $g(\cdot)$ is a known link function. Binary and count observations are two common types of non-Gaussian spatial data, and these can be modeled using the binary SGLMM with logit link and the Poisson SGLMM with log link, respectively. The

general hierarchical framework for spatial generalized linear mixed models is:

$$
\begin{aligned}
&\textbf{Data Model:} && Z(s)|\eta(s) \sim f(\eta(s)) \\
& && \eta(s) = g(\mathbb{E}[Z(s)|\beta, \mathbf{W}), \epsilon(s)]) = X(s)\beta + w(s) + \epsilon(s) \\
&\textbf{Process Model:} && \mathbf{W}|\phi, \sigma^2 \sim N(0, \sigma^2 R_\phi), \qquad \mathbf{W} = \{w(s_1), ..., w(s_n)\} \\
& && \epsilon(s)|\tau^2 \sim N(\mathbf{0}, \tau^2) \\
&\textbf{Parameter Model:} && \beta \sim p(\beta), \phi \sim p(\phi), \sigma^2 \sim p(\sigma^2), \tau^2 \sim p(\tau^2)
\end{aligned}
$$

Other examples of hierarchical spatial models include spatially varying coefficient processes (Gelfand et al., 2003; Mu et al., 2018), covariate measurement error models (Xia and Carlin, 1998; Bernadinelli et al., 1997; Muff et al., 2015), and co-regionalization models for multivariate responses (Banerjee et al., 2014).

1.2.2 Zero-Inflated Spatial Models

Zero-inflated spatial data are spatially dependent observations characterized by an excess of zeros. Observations can be discrete counts or semi-continuous, where the non-zero values are positive real numbers. Zero-inflated spatial observations are commonly encountered in many fields; for instance, counts of harbor seals on glacial ice (Hoef and Jansen, 2007), annual mental health expenditures among US federal employees Neelon et al. (2011), and the number of torrential rainfall events in a region of interestLee and Kim (2017). Standard probability distributions are not sensible for modeling zero-inflated data (cf. Agarwal et al., 2002; Rathbun and Fei, 2006; Lambert, 1992a) as they are unable to account for the large proportion of zeros. Moreover, poor model choice may lead to over- or under-dispersion, where the observed variance is higher or lower, respectively, than the variance of the fitted model.

Two-part models (Mullahy, 1986; Lambert, 1992b) have commonly been used to model zero-inflated spatial data (Agarwal et al., 2002; Hoef and Jansen, 2007; Olsen and Schafer, 2001, .cf). These models generally include two spatial processes: (1) the occurrence process $O(s)$, that specifies the structural zero and non-zero locations; and (2) the prevalence process $P(s)$ that generates positive values (and zeros in some cases) for the non-zero locations. Note that both processes model

spatial dependencies among the observation locations.

Let $Z(s)$ be a zero-inflated observation for spatial location $s \subset \mathcal{D}$ within the spatial domain $\mathcal{D} \in \mathbb{R}^2$. The observation $Z(s)$ are generated as follows:

$$Z(s) = \begin{cases} 0 & \text{if } O(s) = 0 \\ P(s) & \text{if } O(s) = 1. \end{cases}, \tag{1.8}$$

where $O(s)$ and $P(s)$ are the spatial occurrence and prevalence processes, respectively. The occurrence process is specified as $O(s) \sim Bern(\pi(s))$ with spatially varying probabilities $\pi(s) \in (0, 1)$. The prevalence process is modeled as $P(s) \sim f(\theta(s))$ where $f(\theta(s))$ is a discrete or continuous probability distribution with spatially varying model parameters $\theta(s)$. In fact, two-part models are identified by the choice of $f(\cdot)$.

Two-part models for zero-inflated observations typically fall into two classes:

1. **Hurdle Models:** The occurrence process $O(s)$ determines the zero-valued locations. The prevalence process $P(s)$ generates the positive values at the non-zero locations. In the discrete case, $f(\cdot)$ is a zero-truncated distribution such as the zero-truncated Poisson or the zero-truncated negative binomial distribution. For semi-continuous observations, $f(\cdot)$ is a probability distribution with positive support such as a log-normal or gamma distribution.

2. **Mixture Models:** Both the occurrence $O(s)$ and prevalence processes $P(s)$ determine the zero-valued locations. The occurrence process identifies the structural zero-valued locations. The prevalence process $P(s)$ generates values for the structural non-zero-values locations. Here, the prevalence process $P(s)$ generates both zeros and positive values for the non-zero-values locations. In the discrete case, $f(\cdot)$ is a non-degenerate distribution such as the Poisson or Negative-Binomial distribution. For semi-continuous observations, $f(\cdot)$ can be a censored model such as a Tobit Type I.

Both processes, $O(s)$ and $P(s)$, are modeled as spatial generalized linear mixed models (SGLMMs) with the appropriate link functions. The occurrence process $O(s)$ is modeled as a Bernouilli random variable with either a probit or a logit link function. $O(s)$ can also be modeled using a latent probit process (Albert and

Chib, 1993; De Oliveira, 2000), which provides simple full conditional distributions for the random effects. The linear predictor is defined as $\boldsymbol{\eta_o} = X\boldsymbol{\beta_o} + \mathbf{W_o} + \epsilon_o$, where $\mathbf{W_o} \sim \mathcal{N}(0, \sigma_o^2 R_{\phi_o})$ and $\epsilon_o \sim \mathcal{N}(0, \tau_o^2 \mathcal{I})$. Model fitting entails estimating the parameters $\beta_o, \phi_o, \sigma_o^2, \tau_o^2$ as well as the spatial random effects $\mathbf{W_o}$.

The prevalence process $P(s)$ follows a specific probability distribution based on the observation type (counts vs. semi-continuous) and structural assumptions. For HURDLE models, a zero-truncated distribution (e.g., zero-truncated Poisson, zero-truncated negative binomial, lognormal, or gamma) is a sensible choice for $f(\cdot)$. Mixture models utilize a distribution with non-negative support (e.g., Poisson, negative binomial, or Tobit model). Similar to the occurrence process, the prevalence process $P(s)$ is also modeled as an SGLMM with linear predictor $\boldsymbol{\eta_p} = X\boldsymbol{\beta_p} + \mathbf{W_p} + \epsilon_p$, where $\mathbf{W_p} \sim \mathcal{N}(0, \sigma_p^2 R_{\phi_p})$ and $\epsilon_p \sim \mathcal{N}(0, \tau_p^2 \mathcal{I})$. Here, the parameters $\beta_p, \phi_p \sigma_p^2, \tau_p^2$ and spatial random effects $\mathbf{W_p}$ must be estimated. To complete the Bayesian hierarchical framework, prior distributions are specified for the model parameters.

The Bayesian hierarchical framework for two-part models is as follows:

Data Model:	$\mathbf{Z}\|O(s), P(s) \sim \tilde{f}_Z\big(z; O(s), P(s)\big)$
Process Model:	$O(s)\|\pi(s) \sim \text{Bern}(\pi(s))$
	$P(s)\|\theta(s) \sim f(\theta(s))$
Sub-process Model 1:	$\pi(s)\|\eta_o(s) = g_o^{-1}(\eta_o(s))$
(Occurrence)	$\eta_o(s)\|\beta_o, W_o(s), \epsilon_o(s) = X(s)\beta_o + W_o(s) + \epsilon_o(s)$
	$\mathbf{W}_o = \{W_o(s_1), ..., W_o(s_n)\}$
	$\mathbf{W}_o\|\phi_o, \sigma_o^2 \sim \mathcal{N}(\mathbf{0}, \sigma_o^2 R_{\phi_o})$,
	$\epsilon(s)\|\tau_o^2 \sim \mathcal{N}(\mathbf{0}, \tau_o^2)$
Sub-process Model 2:	$\theta(s)\|\eta_p(s) = g_p^{-1}(\eta_p(s))$
(Prevalence)	$\eta_p(s)\|\beta_p, W_p(s), \epsilon_p(s) = X(s)\beta_p + W_p(s) + \epsilon_p(s)$
	$\mathbf{W}_p = \{W_p(s_1), ..., W_p(s_n)\}$
	$\mathbf{W}_p\|\phi_p, \sigma_p^2 \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 R_{\phi_p})$
	$\epsilon(s)\|\tau_p^2 \sim \mathcal{N}(\mathbf{0}, \tau_p^2)$
Parameter Model:	Priors for $\beta_o, \phi_o, \sigma_o^2, \tau_o^2, \beta_p, \phi_p, \sigma_p^2,$ and τ_p^2

where $\tilde{f}_Z\big(z; O(s), P(s)\big)$ is the likelihood function of spatial two-part model. Based on Equation 4.3, $\tilde{f}_Z\big(z; O(s), P(s)\big)$ is defined as:

$$\tilde{f}_Z\big(z; O(s), P(s)\big) = \begin{cases} \pi(s) + (1 - \pi(s)) \times f(0; \theta(s)), & \text{if } z = 0 \\ (1 - \pi(s)) \times f(z; \theta(s)), & \text{if } z > 0. \end{cases}, \tag{1.9}$$

The literature presents alternative distributions for $f(\cdot)$, which can yield richer and more flexible two-part models. For count data, past studies have used the Poisson, negative binomial, zero-truncated Poisson (Lambert, 1992a), translated Poisson (Hoef and Jansen, 2007), zero-truncated negative binomial (Mwalili et al., 2008), generalized Poisson (Gschlößl and Czado, 2008), and binomial distributions (Hall, 2000). In the semi-continuous case, the lognormal distribution may not be appropriate due to the lack of symmetry or fatter tails of the log of the observations. Past studies have used skewed distributions (Dreassi et al., 2014; Liu et al., 2016), t-distributions to model heavy tailed behavior (Neelon et al., 2015), or modeled the prevalence process using scale mixtures of normal distributions (Fruhwirth-Schnatter and Pyne, 2010).

There exists many promising areas of future research for two-part spatial models. For instance, neglecting the cross-correlation between spatial processes can potentially lead to biased inferences (Su et al., 2009). To counteract this, the latent spatial processes, $\mathbf{W}_o$ and $\mathbf{W}_p$, can be modeled jointly (Recta et al., 2012; Oliver, 2003; Neelon et al., 2011; Su et al., 2009); however, this may incur additional computational costs due to large matrix operations (Neelon et al., 2016b). Another open area of research addresses selecting the appropriate type of two-part model (HURDLE or mixture). Past studies (Hu et al., 2011; Wilson, 2015; Xu et al., 2015) have examined the goodness-of-fit between hurdle and mixture models using likelihood ratio tests, AIC, BIC, DIC, or the Vuong test statistic (Vuong, 1989).

However, there is a dearth of research addressing the computational challenges inherent to modeling large zero-inflated spatial data, particularly cases with high-dimensional and heavily correlated latent spatial random effects. One study (Wang et al., 2014) models the presence and abundance of Atlantic cod in 1325 locations along the Gulf of Maine using predictive processes (Banerjee et al., 2008). Other

studies provide methods to facilitate posterior sampling by representing the latent spatial processes as scale mixtures of normal distributions via dirichlet processes (Neelon et al., 2018) or through Pólya-Gamma mixtures (Neelon et al., 2019).

1.2.3 Basis Representation of Spatial Random Fields

Hierarchical spatial models are subject to computational challenges borne out of the heavily correlated and high-dimensional spatial random effects $\mathbf{W}$. Model fitting typically requires a costly evaluation of an $n-$dimensional multivariate normal likelihood function ($\mathcal{O}(n^3)$) at each iteration of the MCMC algorithm. In addition, heavily correlated spatial random effects often leads to poor mixing in MCMC algorithms (cf. Christensen et al., 2006; Haran et al., 2003). In the literature, basis representation approaches (Cressie and Wikle, 2015, cf.) have been used to address these challenges. In this framework, the spatial random effects $\mathbf{W} = (W(s_1), W(s_2), ..., W(s_n))$ as a linear combination of basis functions:

$$\mathbf{W} \approx \mathbf{\Phi}\delta \qquad , \qquad \delta \sim \mathcal{N}(0, \Sigma_\delta(\xi)),$$

where $\mathbf{\Phi}$ is an $n \times p$ basis function matrix where each column denotes a basis function, $\delta \in \mathbb{R}^p$ are the basis coefficients, and $\Sigma_\delta(\xi)$ is the $p \times p$ covariance matrix for the coefficients with covariance parameter ξ. Under this setting, the basis functions are a collection of distinct spatial patterns used to construct the latent spatial random field. Basis representation approaches can reduce computational costs by bypassing large matrix operations such as matrix inversions (Banerjee et al., 2008; Higdon, 1998), reducing the dimensionality of latent spatial random effects (Guan and Haran, 2018; Hughes and Haran, 2013), inducing sparse matrix operations via localized basis functions (Katzfuss, 2017; Nychka et al., 2015; Cressie and Johannesson, 2008; Lindgren et al., 2011), and reducing correlation among the latent spatial random field (Christensen et al., 2006).

Spatial basis functions readily extends to the Bayesian hierarchical framework. To illustrate, basis functions can be embedded into the Bayesian hierarchical frame-

work for Gaussian spatial data (Equation 1.7) as so:

$$
\begin{aligned}
&\textbf{Data Model:} && Z(s)|\beta, \mathbf{W}, \tau^2 \sim N(X(s)\beta + w(s), \tau^2) \\
&\textbf{Process Model:} && \mathbf{W} \approx \mathbf{\Phi}\delta, \qquad \mathbf{W} = \{w(s_1), ..., w(s_n)\} \\
& && \delta \sim \mathcal{N}(0, \Sigma_\delta(\xi)) \\
&\textbf{Parameter Model:} && \beta \sim p(\beta), \xi \sim p(\xi), \tau^2 \sim p(\tau^2)
\end{aligned}
$$

Selecting the appropriate set of basis functions $\mathbf{\Phi}$ remains an open area of research. In the literature, there exists a wide array of spatial basis functions such as: (1) bi-square (radial) basis functions (Cressie and Johannesson, 2008) with varying resolutions (Katzfuss, 2017; Nychka et al., 2015); (2) empirical orthogonal functions (Cressie and Wikle, 2015), or spatial representations of principal component analysis; (3) predictive process basis functions (Banerjee et al., 2008); (4) Moran's basis functions for spatial eigenfiltering (Hughes and Haran, 2013; Griffith, 2003); (5) piecewise Linear functions on a triangulation of the spatial domain (Lindgren et al., 2011); (6) square roots of the correlation matrix of the spatial random effects via cholesky factorization Christensen et al. (2006) or approximate eigendecompositions (Banerjee et al., 2013; Guan and Haran, 2018); (7) W-wavelets(Shi and Cressie, 2007) and multiresolution wavelet basis functions (Nychka et al., 2002); (8) Fourier basis functions composed of sine and cosine curves (Royle and Wikle, 2005); and (9) Gaussian kernel basis functions (Higdon, 1998). These methods tend to be computationally efficient as they bypass large matrix operations and in some cases de-correlate and reduce the dimensions of the spatial random effects $\mathbf{W}$.

Many of these basis functions (radial, Fourier, wavelets, Gaussian kernels) rely on an overcomplete set of basis functions where the number of basis functions are much larger than the number of observations. Consequently, this can potentially increase computational costs and the selection of the resolutions are important. Low rank methods, such as predictive processes or random projections, still require large matrix operations, albeit smaller than the gold standard.

In Chapter 3, I propose a fast and extendable approach to fit hierarchical spatial models. This approach approximates a continuous spatial process using a discretized Gauss-Markov random field and a data-driven set of basis func-

tions. I addresses the high-dimensionality of the spatial random effects by using a projection-based intrinsic autoregression approach which simultaneously reduces the dimensions and de-correlates the spatial random effects. Moreover, I provide an automated heuristic to select the appropriate number of basis functions. In Chapter 4, I extend the PICAR approach to zero-inflated spatial models to address key computational challenges and also provide practical guidelines for model selection.

1.3 Summary of Contributions

This dissertation makes the following contributions:

- I provide a fast particle-based approach for calibrating complex computer models, namely the Pennsylvania 3D Antarctic Ice Sheet model (PSU3D-ICE). This novel calibration method is designed to calibrate complex computer models with moderate single model run times (5 seconds to 15 minutes) and a moderate number of model parameters (5 to 15). This work is co-authored by Murali Haran, Rob Fuller, David Pollard, and Klaus Keller, and the corresponding manuscript has been accepted for publication by the *Annals of Applied Statistics*.

- I conduct a formal investigation of how the choice of calibration methods impacts sea level rise projections. I compare three calibration approaches for the PSU3D-ICE Antarctic ice sheet model: (1) particle-based approach with 11 parameters; (2) emulation-calibration with three parameters; and (3) emulation-calibration with 11 parameters. I find that (2) and (3) either drastically underestimates the tail-area risk for sea level rise projections or provides highly inaccurate projections.

- I developed a computationally efficient and extendable method (PICAR) to fit high-dimensional hierarchical spatial models. This approach discretizes the continuous latent spatial random field and employs a basis representation the underlying spatial process. The method is scalable to high-dimensional datasets and extends to a wide array of spatial hierarchical spatial models (e.g. spatial generalized linear mixed models, spatially varying coefficients

models, spatial ordered categorical models). Moreover, PICAR is readily extendable to popular programming platforms such as `stan` and `nimble`. This manuscript is co-authored with Murali Haran and is under revision.

- I propose a scalable method for modeling high-dimensional zero-inflated spatial observations. This approach extends the PICAR representation of latent spatial processes to zero-inflated spatial models. I demonstrate this computationally efficient approach on several simulated and real-world datasets. A manuscript based on this work (co-authored with Murali Haran) is currently in preparation.

1.4 Thesis Organization

The remainder of this dissertation is organized as follows. In Chapter 2, I describe the challenges in projecting the long-term behavior of the Antarctic ice sheet and its effect on global sea level rise. Next, I present a fast particle-based approach for calibrating a state-of-the-art three-dimensional computer model of the Antarctic ice sheet (PSU3D-ICE). In Chapter 3, I provide an overview of hierarchical spatial models as well as a discussion of the associated computational challenges. Then, I propose an extendable projection intrinsic conditional autoregression (PICAR) approach for fitting hierarchical spatial models and demonstrate this method on simulated and real-data examples. In Chapter 4, I introduce spatial two-part models for zero-inflated spatial data and discuss the underlying inferential and computational challenges. Then, I propose a computationally efficient approach for fitting high-dimensional zero-inflated spatial observations. Finally, in Chapter 5, I summarize my contributions and discuss avenues for future research.

Chapter 2

A Fast Particle-Based Approach for Calibrating a 3-D Model of the Antarctic Ice Sheet

In this section, I present a fast particle-based approach for computer model calibration with applications to a 3-D Model of the Antarctic Ice Sheet. I propose a sequential Monte Carlo-based calibration method that drastically reduces calibration wall times while still preserving close approximations. This method enables important computer experiments, which were computationally infeasible under existing approaches. This chapter is published as a manuscript (Lee et al., 2020). All authors co-designed the overall study. BSL and MH formulated the statistical method. BSL wrote the computer code for calibration, designed the Pliocene window analysis, and wrote the first draft of the manuscript. RF integrated the calibration method into the Cheyenne high performance computing system. MH edited the text. KK designed the comparative methods analysis and edited the text. DP provided code and data for the 80 km resolution PSU3D-ICE model, designed the prior sensitivity study, and edited the text.

2.1 Introduction

How much will the Antarctic ice sheet contribute to future sea level rise? The geological records suggest that ice sheets can quickly contribute considerable amounts to global sea level rise (Deschamps et al., 2012), in some cases up to 58 m (Fretwell et al., 2012). Projections of future sea level rise depend on deeply uncertain projections of the Antarctic ice sheet's (AIS) mass loss (Le Bars et al., 2017; Wong et al., 2017; Le Cozannet et al., 2017). Close to eight percent of the current global population is threatened by a five meter rise in sea level (Nicholls et al., 2008) and 13 percent of the global urban population is threatened by a ten meter sea level rise (McGranahan et al., 2007). Quantifying and characterizing the long-term behavior of the Antarctic ice sheet is hence a key input to the design of coastal risk management strategies (cf. Garner and Keller, 2018; Sriver et al., 2018; Oppenheimer and Alley, 2016).

Ice sheet models rely on poorly constrained parameters, and recent studies show that uncertainty in model parameters results in highly uncertain projections of sea level change (Stone et al., 2010; Applegate et al., 2012; Fitzgerald et al., 2012; Collins, 2007); thereby affecting climate risk decision-making (O'Neill et al., 2006; Hannart et al., 2013). Recent studies have addressed parametric uncertainty via calibration studies using modern observations, but these are either limited to simple ice sheet models (Ruckert et al., 2017; Fuller et al., 2017) or a small number of model parameters (Chang et al., 2016b; Edwards et al., 2019; Schlegel et al., 2018). Numeric solvers have been used to infer the field of basal sliding parameters from satellite observations (Isaac et al., 2015b,a).

Ice sheet models vary in complexity, and the key drivers of computational cost are the spatial and temporal resolutions. Simpler models (cf. Shaffer, 2014; Bakker et al., 2016) have short computer model run times on the order of a few seconds, but they may oversimplify or even exclude important physical processes. More complex models (cf. DeConto and Pollard, 2016; Larour et al., 2012; Greve, 1997; Rutt et al., 2009) can better represent key ice dynamics and typically run at higher spatio-temporal resolutions. However, they require longer model run times. Here, we use a relatively complex ice sheet model, the Pennsylvania State University 3D ice sheet model (PSU3D-ICE) (Pollard and DeConto, 2012a), but

with considerably coarser resolution than in previous work, so that each set of simulations for this study takes on the order of 10 to 15 minutes of wall time.

Past studies calibrate simpler models with many model parameters via Markov Chain Monte Carlo (MCMC) (cf. Ruckert et al., 2017; Bakker et al., 2016; Petra et al., 2014); these approaches are effective in the context of computationally very inexpensive models (model run times of a few seconds), and hence do not extend to the kind of models we consider in this manuscript. Some studies have employed emulation-calibration methods (Sansó et al., 2008; Liu et al., 2009a; Bhat et al., 2010) to calibrate computer models with long run times, but these approaches are applicable to only a small number of parameters. For computer models with longer run times and a large number of model parameters, emulation-calibration can be computationally prohibitive because building an accurate emulator requires a very large set of training data (Bastos and O'Hagan, 2009; Maniyar et al., 2007).

We propose calibrating an ice sheet model which (1) accounts for important physical processes; (2) includes several key parameters to analyze and quantify parametric uncertainty; and (3) expands the calibration dataset to the Pliocene. For this study, the Antarctic ice sheet model runs at a spatial resolution of 80 km and temporal resolution of eight years, which is a compromise between preserving reasonable accuracy of physical simulations versus maintaining a feasible model run time. We estimate that current rigorous methods for calibrating this model via Markov chain Monte Carlo would take roughly on the order of years of wall time. We investigate methods for calibration that are amenable to heavy parallelization and computationally efficient, thereby reducing the computational wall time from years to hours. We find that these methods are broadly applicable to computer models with a moderate model run time (6 seconds to 15 minutes) and a moderate number of model parameters (5 to 20), based on available computing resources. While this does not cover more complex models or larger number of parameters, our methods are applicable to many scientifically important and policy-relevant computer models.

Studying the Antarctic ice sheet's future behavior motivates the need for a computationally efficient approach for computer model calibration. We turn to sequential Monte Carlo methods (cf. Doucet et al., 2000; Del Moral et al., 2006; Chopin, 2002), building upon particle-based methods for computer model calibra-

tion (Higdon et al., 2008; Kalyanaraman et al., 2016). Our approach builds upon an adaptive tempering schedule and an adaptive mutation stage (Jasra et al., 2011), which have been used for Bayesian variable selection (Schäfer and Chopin, 2013), Bayesian model comparison (Zhou et al., 2016), and estimating initial conditions of the Navier-Stokes system of equations (Kantas et al., 2014; Llopis et al., 2018).

By using massive parallelization in a high performance computing environment, we obtain a dramatic speed-up over current MCMC-based calibration methods, roughly reducing wall time by a factor of 3000. We also limit expensive computer model runs by imposing stopping rules and adaptive sampling techniques. We provide practical guidelines designed to: (1) reduce total wall time; (2) limit the number of expensive computer model runs; and (3) simplify implementation for the user. Our computationally efficient calibration approach is readily applicable to many computer models for which rigorous calibration may be currently infeasible.

We note that we focus on a 'static' system where all observations are available at once; hence, there is only one posterior distribution of interest, which we approximate using our particle-based approach. The PSU3D-ICE model is dissipative where it evolves to a single constant steady state for a given set of parameter values and external forcing (Willems, 1972). Unlike choatic systems such as global weather models, "microscopic" changes in the initial states do not change the results; in other words, there is no "butterfly effect" (Lorenz, 1972). We use our approach to calibrate the PSU3D-ICE model (DeConto and Pollard, 2016) using paleoclimate data and modern observational records. Previous work focuses on calibrating the PSU3D-ICE model using fewer parameters (Chang et al., 2016b; Edwards et al., 2019) or surrogate models using limited training data (Chang et al., 2016a). Using our new method, we show that the information regarding the extent of the Antarctic ice sheet in the Pliocene era strongly influences parametric and projection uncertainty. We find that using improved geological data and analysis to characterize the Antarctic ice sheet's contribution to sea level rise in the Pliocene can bring about considerably sharper sea level projections for future centuries.

The paper is structured as follows. In Section 2.2, we provide an overview of the ice sheet model (PSU3D-ICE). In Section 2.3, we describe the model calibration framework and discuss challenges with current calibration methods. We propose our fast particle-based approach for computer model calibration in Section 2.4. In

Section 2.5, we demonstrate the application of our method to a simulated example. In Section 2.6 we apply our method to the PSU3D-ICE model and report our scientific conclusions. We end with caveats and directions for future research in Section 2.7.

2.2 Description of computer model and data

In this section, we provide background information for the PSU3D-ICE Antarctic ice sheet model (DeConto and Pollard, 2016) as well as the paleoclimate records and modern observations used to calibrate the model.

2.2.1 The PSU3D-ICE model

The PSU3D-ICE model simulates the long-term dynamics of continental ice sheets. It has previously been applied to past and future variations of the Antarctic ice sheet (Pollard and DeConto, 2009, 2012a; Pollard et al., 2015, 2016, 2017). Slow ice deformation under its own weight is modeled by scaled dynamical equations for internal shear, horizontal stretching, and basal sliding. Other variables and processes include internal ice temperatures, bedrock deformation beneath the ice load, surface snowfall and melting, oceanic melting beneath floating ice shelves, and calving of ice into the ocean (Pollard and DeConto, 2012a). A recently proposed mechanism called Marine Ice Cliff Instability (MICI) that can drastically attack ice in marine basins, involving hydrofracturing due to surface liquid water and structural failure of tall ice cliffs, is included here (Pollard et al., 2015; DeConto and Pollard, 2016). Note that this mechanism has recently been questioned (Edwards et al., 2019; Golledge et al., 2019).

For the simulations in this study, a polar stereographic grid spanning Antarctica is used with a horizontal resolution of 80 kilometers (km), which yields a model run time of approximately 10 to 15 minutes for each set of past and future simulations described below. This is a considerably coarser spatial resolution than previous continental-scale applications, which have used resolutions of 10 to 40 km. However, sensitivity tests with the model show reasonable independence of results with model resolution, due to the grid-independent parameterization of important

sub-grid processes such as grounding-line flux and cliff failure (Pollard et al., 2015). Those tests and the reasonable agreement in additional limited offline tests at 80 km vs. finer resolutions indicate that the coarser resolution is adequate for this study.

We evaluate the PSU3D-ICE model over three separate time periods. As in previous ensemble work with this model (Chang et al., 2016a,b; Pollard et al., 2016; DeConto and Pollard, 2016), the time periods are selected to include major ice-sheet variations that stringently test the model and have at least some paleo data to provide useful quantitative constraints. The three time periods are: (1) a single episode of high sea level rise during the warm mid-Pliocene (which extended roughly from 3.2 to 2.6 million years before present); (2) the Last Interglacial period around 125,000 to 115,000 years ago, at the start of the last Pleistocene glacial-interglacial cycle when global climate was slightly warmer than today, the major Northern Hemispheric ice sheets were most recently absent prior to the modern interglacial period, Greenland was smaller, and the West Antarctic ice sheet may have undergone major collapse; and (3) the last deglacial period from the Last Glacial Maximum about 20,000 years ago to the present, and then 5,000 years into a warmer future. In Figure 2.1, we present 1500 model simulations from the PSU3D-ICE model for all three time periods as well as projections until year 2500. We describe the three model simulations below.

To represent a single high sea level episode during the warm mid-Pliocene era (roughly 3.2 to 2.6 million years before present), we initialize the ice sheet model to modern conditions and run the model forward for 5,000 years. As described in previous Pliocene applications (Pollard et al., 2015, 2017), atmospheric climatic forcing is provided by the RegCM3 regional climate model (Pal et al., 2007) adapted for polar regions, driven by the GENESIS v3 global climate model (Alder et al., 2011). The atmospheric carbon dioxide concentration is set to at 400 parts per million by volume (ppmv), and a warm austral summer orbit is specified. We use oceanic temperatures from the modern World Ocean Atlas database (Levitus et al., 2012), with a +2 °C uniform perturbation added to represent mid-Pliocene ocean warming. Atmospheric monthly cycles of surface air temperature and precipitation are used to compute melting and annual mass balance on the ice-sheet surface, and oceanic temperatures are used to compute basal melting under floating ice shelves

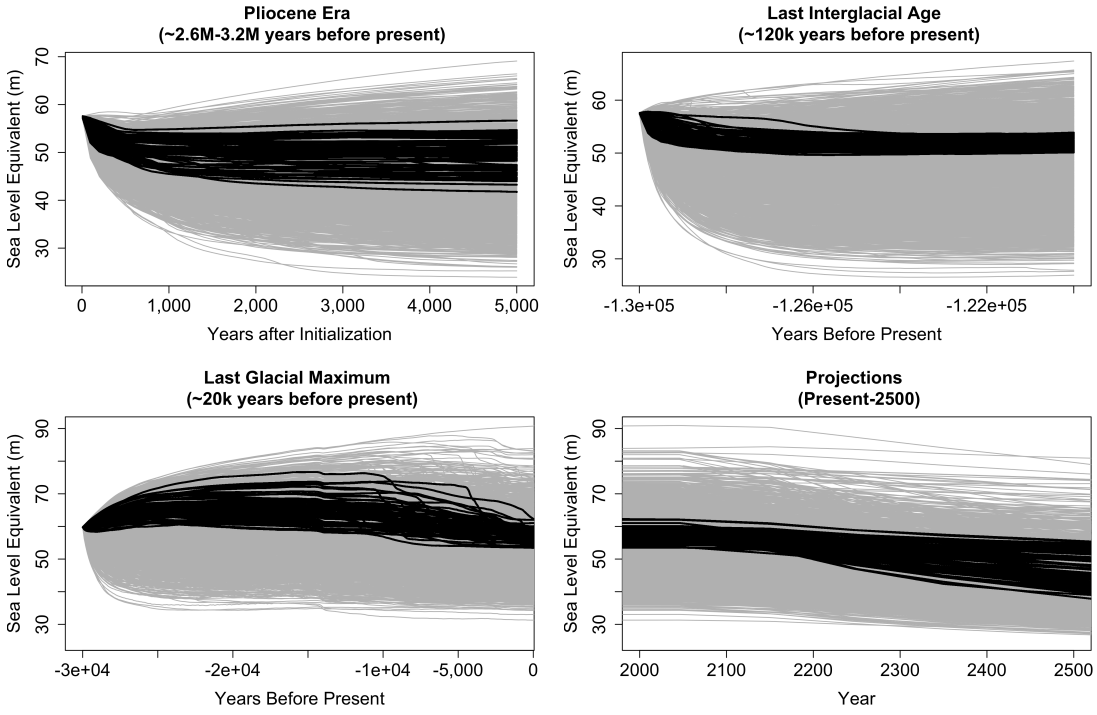

Figure 2.1. Time series of 1500 simulated model output from the PSU3D-ICE model where each model run corresponds to a line. Data are generated using 1500 parameter sets from the prior distribution. The y-axis denotes the Antarctic ice sheet's contribution to sea level change in meters (m). We approximate the present as year 1950. Model simulations that have a non-zero likelihood are denoted by black lines and runs that have a zero likelihood are displayed in light gray. (Top left) Model output for the Pliocene era model run where the x-axis denotes years after initialization. (Top right) Model output for the Last Interglacial Age where the x-axis denotes years before the present. (Bottom left) Model output for the Last Glacial Maximum where the x-axis denotes years before the present. (Bottom right) Model projections for 2000-2500 where the x-axis represents years.

(Pollard et al., 2015).

For the Last Interglacial (LIG), we initialize the ice sheet model to modern conditions and run the model from 130,000 to 120,000 years before present (130 ka to 120 ka). As described in DeConto and Pollard (2016), LIG climates are specified as uniform perturbations to modern climatology (Le Brocq et al. (2010) for atmosphere, and Levitus et al. (2012) for ocean). The atmospheric and ocean temperature perturbations vary step-wise in time. From 130 ka to 125 ka, they are +1.97°C and +1.70°C respectively. From 125 ka to 120 ka, they are +1.41°C and +1.51°C respectively.

The Last Glacial Maximum, modern, and future eras are simulated in one continuous run, over the last 40,000 years through the last deglacial period to modern, and extended 5,000 years into the future. As described in Pollard et al. (2016), the model is initialized appropriately at 40 ka (40,000 years Before Present or BP, relative to 1950 AD) from a previous long-term run. Atmospheric forcing is supplied using a modern climatological Antarctic dataset (Le Brocq et al., 2010), with uniform cooling perturbations applied proportional to a deep sea-core δ^{18}O record (Pollard and DeConto, 2009, 2012a). Oceanic forcing is supplied from a coupled Atmosphere-Ocean General Circulation Model (AOGCM) simulation of the last 20,000 years (Liu et al., 2009b). After reaching present day, each run is extended for 5,000 years with atmospheric and oceanic forcing as described in DeConto and Pollard (2016), for the Representative Concentration Pathway (RCP) 8.5 scenario of future greenhouse gas emissions and concentrations (Meinshausen et al., 2011), often called 'business as usual'. Atmospheric temperatures and precipitation are obtained by appropriately weighting previously saved simulations of the RegCM3 regional climate model for particular carbon dioxide levels, and oceanic temperatures are supplied from an archived transient NCAR global model simulation (Shields and Kiehl, 2016).

After each model run, we extract the pertinent model output, specifically the Antarctic ice sheet's contribution to sea level change (m), total ice volume (km^3), and total grounded ice area (km^2). We then compare this to the corresponding paleo- or modern observational records. In this study, we examine 11 model parameters considered to be important in modeling the behavior of the Antarctic ice sheet - OCFACMULT, OCFACMULTASE, CRHSHELF, CRHFAC, ENHANCESHEET, ENHANCESHELF, FACEME-LTRATE, TAUASTH, CLIFFVMAX, CALVLIQ, and CALVNICK. Detailed descriptions of each parameter are provided in the Supplement (Lee et al., 2019).

We note that this is a much larger number of parameters than typically considered for models with such detailed dynamics. The ice sheet model has many more parameters than the 11 chosen here. The values for many of them are reasonably well established in the glaciological literature, resulting from published work over the last several decades applying similar models to the Antarctic ice sheet. Those parameters mostly involve terrestrial processes (i.e., where ice is grounded

on bedrock) that are constrained directly or indirectly by observational data of the modern ice sheet, and/or laboratory ice physics, such as the rheology of ice, ice streaming vs. shearing flow, basal sliding coefficients, and modern ice distribution and thicknesses. The 11 parameters chosen here can have large effects on the results, but are not well constrained by modern observations because they apply to processes (1) that have occurred in the past and expected in the future, but are not active today, or (2) are undergoing rapid change in recent decades. Examples of (1) are basal sliding coefficients for bedrock in modern ocean regions where grounded ice advanced during past glacial maxima, and the timescale of bedrock rebound under varying ice loads. Examples of (2) are coefficients for oceanic melting at the base of floating ice shelves, and oceanic melting at vertical ice fronts. A subset of these parameters have been used in more limited ensembles with this model (Chang et al., 2016b,a; Pollard et al., 2016, 2017), but here the 11 parameters constitute the bulk of important yet relatively unconstrained parameters in the model.

2.2.2 Paleoclimate records and modern observations

For the paleoclimate records, we use the Antarctic ice sheet's contribution to sea level change in the following eras: Pliocene (∼2.6-3.2 million years before present); the Last Interglacial Age (∼125,000 to 115,000 years before present); and the Last Glacial Maximum (∼20,000 years before present). We specify the Antarctic ice sheet's contribution to sea level change in terms of global mean sea level equivalents (SLE) relative to the modern ice sheet, thereby correctly allowing for marine ice grounded below sea level. The base units are meters (m). We adopt the following ranges for the paleoclimate records, which account for considerable uncertainty in published estimates (cf. Kopp et al., 2009; Dutton et al., 2015): (1) 5 m to 25 m for the Pliocene (Naish et al., 2009; Rovere et al., 2014; Cook et al., 2013); (2) 3.5 m to 7.5 m for the Last Interglacial Age (Fuller et al., 2017; DeConto and Pollard, 2016); and (3) -5 m to -15 m for the Last Glacial Maximum (Ruckert et al., 2017; Pollard et al., 2016).

Modern observations include total volume and grounded area of the Antarctic ice sheet, as well as ten spatial locations that currently have ice present.

Units for total volume and total grounded ice area are cubic kilometers (km^3) and square kilometers (km^2) respectively. Observations come from the Bedmap2 dataset (Fretwell et al., 2012), which provide the most recent gridded maps of ice surface elevation, bedrock elevation, and ice thickness. The Bedmap2 maps are generated using multiple sources, including satellite altimetry, airborne and ground radar surveys, and seismic sounding.

2.3 Model calibration framework

In this section, we describe the general computer model calibration framework. In computer model calibration, key computer model parameters are estimated by comparing the computer model output and observational data (cf. Chang et al., 2016a; Kennedy and O'Hagan, 2001; Bayarri et al., 2007; Bhat et al., 2010). Calibration methods also account for key sources of uncertainty such as model-observation discrepancy and observational error (Kennedy and O'Hagan, 2001; Bayarri et al., 2007; Brynjarsdottir and O'Hagan, 2014). We describe a model for output in the form of spatial data as this directly relates to our simulated data example in Section 2.5; a time series version of this applies to the PSU3D-ICE model in Section 2.6.

Let $Y(s, \theta)$ be the computer model output at the spatial location $s \in \mathcal{S} \subseteq \mathbb{R}^2$ and the parameter setting $\theta \in \Theta \subseteq \mathbb{R}^d$. $\mathcal{S}$ is the spatial domain of the process, and Θ is the parameter space of the computer model with integer d being the number of input parameters. $\mathbf{Y} = (Y(s_1, \theta_i), ..., Y(s_n, \theta_i))^T$ is the computer model output at parameter setting θ_i and spatial locations $(s_1, ..., s_n)$. $\mathbf{Z} = (Z(s_1), ..., Z(s_n))^T$ is the observed spatial process at locations $(s_1, ..., s_n)$.

We model the observational data Z as follows,

$$Z = Y(\theta) + \delta + \epsilon, \tag{2.1}$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$ is independently and identically distributed observational error, and δ is a systemic data-model discrepancy term. The discrepancy δ is modeled as a zero-mean Gaussian process, where $\delta \sim N(0, \Sigma_\delta(\xi_\delta))$. This discrepancy term is essential for parameter calibration (Bhat et al., 2010; Bayarri et al., 2007) and

ignoring it may yield biased and overconfident estimates and projections (Brynjarsdottir and O'Hagan, 2014). $\Sigma_\delta(\xi_\delta)$ is the spatial covariance matrix between spatial points $s_1, ..., s_n$ with covariance parameters ξ_δ. We set standard prior distributions for the model parameters, θ, and observational error variance, σ^2_ϵ. On the other hand, informative priors are necessary for the discrepancy term's covariance parameters ξ_δ. Then, we infer θ, σ^2, and ξ_δ by sampling from the posterior distribution, $\pi(\theta, \sigma^2_\epsilon, \xi_\delta|Z)$, via Markov Chain Monte Carlo (MCMC).

Challenges with computer Model Calibration

We focus on a specific class of computer models, characterized by (1) a moderate run time (6 seconds to 15 minutes); and (2) moderately large parameter space (5 to 20 parameters). The modified PSU3D-ICE Antarctic ice sheet model (Section 2.2) fits the specifications for this class of computer models. Several other important models that can potentially be modified to fit within this class are single column atmospheric models (Bony and Emanuel, 2001; Dal Gesso and Neggers, 2018; Gettelman et al., 2019), hydrological soil moisture models (Sorooshian et al., 1993; Liang et al., 1994), simplified earth systems models (Monier et al., 2013), and integrated multi-Sector models for human and earth dynamics (Kim et al., 2006).

The calibration framework requires running the computer model once for each iteration of the MCMC algorithm. Subject to overall calibration wall times, MCMC-based calibration methods are well suited to computer models that run very quickly, typically under 6 seconds per model run. The PSU3D-ICE model takes approximately 10 to 15 minutes per run on a single 2.3-GHz Intel Xeon E5-2697V4 (Broadwell) processor. We estimate that a standard MCMC-based calibration approach for this would take on the order of 2.9 years to approximate the posterior distribution $\pi(\theta|Z)$.

Surrogate methods such as Gaussian process-based emulators are well suited to computer models with long run times. A good design is important for building accurate surrogates. Dense sampling schemes, such as full factorial or fractional factorial designs, capture higher order interactions; however, running the computer model at each of the design points is costly. Space-filling designs such as the Latin

Hypercube Design (McKay et al., 2000; Steinberg and Lin, 2006; Stein, 1987) or adaptive experimental designs (Chang et al., 2016a; Gramacy and Apley, 2015; Urban and Fricker, 2010; Queipo et al., 2005) use fewer design points, but may possibly generate low-fidelity surrogate models by ignoring higher order interactions among inputs (Liu and Guillas, 2017). Since the PSU3D-ICE model exhibits non-linear dependencies among input parameters (Pollard and DeConto, 2012a), we would be limited to 6 or fewer parameters using standard emulation-calibration techniques (with our available computing resources).

2.4 Fast particle-based calibration

In this section, we present a fast particle-based method to calibrate computers models with moderate model run time (6 seconds to 15 minutes) and a moderate number of model parameters (5 to 20). We begin with a description of a sequential sampling-importance-resampling algorithm. Then, we present modifications to the algorithm designed to improve computational efficiency. We examine advantages and limitations of our approach. Finally, we discuss tuning mechanisms for our method and provide practical guidelines.

2.4.1 Sequential sampling-importance-resampling with mutation

We propose a series of sampling-importance-resampling with mutation operations, which includes evolving importance and target distributions. The objective is to efficiently approximate a target distribution using a swarm of evolving particles. Our approach falls under the umbrella of Sequential Monte Carlo algorithms (Del Moral et al., 2006; Doucet et al., 2000; Liu and West, 2001), which have gained wide practical use (cf. Kantas et al., 2015; Papaioannou et al., 2016; Kalyanaraman et al., 2016; Jeremiah et al., 2011; Morzfeld et al., 2018). In particular, we build upon the Iterated Batch Importance Sampling (IBIS) (Chopin, 2002; Crisan and Doucet, 2000) method.

Sampling-importance-resampling

Sampling-Importance-Resampling (Gordon et al., 1993; Doucet et al., 2001) is a sampling method used to approximate a target distribution $\pi(\theta)$ using samples from an importance distribution $q(\theta)$. Suppose we want to estimate $\mu = E_\pi\big[g(\theta)\big]$. Given $q(\theta) > 0$ whenever $g(\theta)\pi(\theta) > 0, \quad \forall\theta \in \Theta$, we observe that $E_\pi\big[g(\theta)\big] = E_q\Big[g(\theta)w(\theta)\Big]$, where $w(\theta) = \frac{\pi(\theta)}{q(\theta)}$ is the importance weight and $\sum_{i=1}^{N} w(\theta_i) = 1$. The importance sampling estimator is $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{N} g(\theta_i)w(\theta_i)$ and $\hat{\mu}_n \to \mu$ with probability 1 by the strong law of large numbers. For target distributions with an unknown normalizing constant, such as the posterior distribution of the model calibration parameters $\pi(\theta|Z)$, the importance weights $w(\theta_i)$, must be normalized.

Sampling-importance-resampling extends the ideas of importance sampling to generate an approximation of a target distribution via samples from an importance distribution and corresponding importance weights (Gordon et al., 1993). The target distribution $\pi(\theta)$, is approximated by the empirical distribution of the samples $\hat{\pi}(\theta)$, and their corresponding normalized weights $\tilde{w}(\theta_i)$'s:

$$\pi(\theta) \approx \hat{\pi}(\theta) = \sum_{i=1}^{N} \tilde{w}(\theta_i)\delta(\theta_i),$$

where $\delta(\theta_i)$ is the Dirac measure that puts unit mass at θ_i and $\sum_{i=1}^{N} \tilde{w}(\theta_i) = 1$.

Poor choices of importance distributions may yield inaccurate approximations of the target distribution (Doucet et al., 2000) due to weight degeneracy and sample impoverishment. As a result, the bulk of the resampled particles, θ_i, do not reside in the high-probability regions of $\pi(\theta)$. Weight degeneracy occurs when almost all of the samples drawn the importance function have near-zero importance weights leaving just a few samples with any significant weights. Multinomial resampling using the normalized importance weights $\tilde{w}(\theta_i)$ can combat weight degeneracy by eliminating the particles with very small important weights and replicating those with higher weights (Gordon et al., 1993; Doucet et al., 2000). After re-sampling, we reset all of the importance weights $w(\theta_i)$ to $1/N$ and replace the weighted empirical distribution $\hat{\pi}(\theta)$ with an unweighted empirical distribution $\ddot{\pi}(\theta)$:

$$\ddot{\pi}(\theta) = \frac{1}{N}\sum_{i=1}^{N} N_i\delta(\theta_i),$$

where N_i is the number of replicates corresponding to particle θ_i and $\sum_{i=1}^{N} N_i = N$.

Weight degeneracy can lead to sample impoverishment where a small subset of particles θ_i's are heavily replicated in the re-sampling step; hence, few unique particle remain. The unweighted/re-sampled empirical distribution $\ddot{\pi}(\theta)$ may poorly approximate the true target distribution $\pi(\theta)$. To alleviate sample impoverishment, mixture approximations (Gordon et al., 1993) or kernel smoothing methods (Liu and West, 2001) can mutate or rejuvenate the replicated particles. However, these methods may not scale well to high-dimensional target distributions (Doucet et al., 2000).

An alternative method mutates the replicated particles with samples from $K(\theta_i^{(t-1)})$, the Metropolis-Hastings transition kernel (Gilks and Berzuini, 2001), whose stationary distribution is also the target distribution $\pi(\theta)$. Here we run J Metropolis-Hastings updates for each particle θ_i, for $i = 1, ..., N$. Other mutation schemes use genetic algorithms (Zhu et al., 2018) or different transition kernels, $K(\cdot)$ (Papaioannou et al., 2016; Murray et al., 2016). The length of the Markov chain, J, will be short and dependent on computing resources. We set the jth sample drawn via MCMC as the mutated particle $\tilde{\theta}_i$. Since $\tilde{\theta}_i \sim \pi(\theta)$, the resulting empirical distribution $\tilde{\pi}(\theta)$ approximates the target distribution $\pi(\theta)$:

$$\pi(\theta) \approx \tilde{\pi}(\theta) = \sum_{i=1}^{N} \tilde{\theta}_i \delta(\tilde{\theta}_i).$$

Even with the mutation step, sampling-importance-resampling may incur large computational costs. Poor choices of importance distributions may result in extreme sample impoverishment, due to the large discrepancy between the importance and target distribution. Here, the mutation stage typically requires very long (and costly) chains of the Metropolis-Hastings algorithm to move the particles into the high-probability regions of the target distribution (Li et al., 2014).

Sequential sampling-importance-resampling

Our fast particle-based approach addresses the limitations noted above. We propose a series of intermediate posterior distributions $\pi_t(\theta|Z)$, for $t = 1, ..., T$ which will act as importance and target distributions. Existing methods use intermediate

posterior distributions for parameter estimation of static systems (Chopin, 2002; Papaioannou et al., 2016; Nguyen, 2014), uncertainty quantification for chemical processes (Kalyanaraman et al., 2016), and calculating maximum-likelihood estimates for hierarchical systems (Lele et al., 2007).

Intermediate posterior distributions can be generated using likelihood tempering (Chopin, 2002; Neal, 2001; Liang and Wong, 2001). For each intermediate posterior distribution $\pi_t(\theta|Z)$, the likelihood component is a fractional power of the original likelihood $p(Z|\theta)$. The tth intermediate posterior distribution, $\pi_t(\theta)$, is generated as follows:

$$\pi_t(\theta|Z) \propto p(Z|\theta)^{\gamma_t} p(\theta), \tag{2.2}$$

where γ_t's are determined according to a schedule where $\gamma_0 = 0 < \gamma_1 < \cdots < \gamma_T = 1$.

For cycle $t = 1$, we set the importance distribution to be the prior distribution $p(\theta)$, and the target distribution to be the first intermediate posterior distribution, $\pi_1(\theta|Z)$. For cycle t, the importance distribution is $\pi_{t-1}(\theta|Z)$ and the target distribution is $\pi_t(\theta|Z)$. Note that the likelihood incorporation schedule need not be uniform. For instance, more of the likelihood can be incorporated into the earlier intermediate posterior distributions.

Finally, we mutate the particles via short runs of the Metropolis-Hastings algorithm, where the stationary distribution is $\pi_t(\theta)$, the tth intermediate posterior distribution. Note that the importance and target distributions are consecutive (tth and $t + 1$th) intermediate posterior distributions, so there is considerable overlap between the high-probability regions of the two distributions. Convergence results for this family of Sequential Monte Carlo algorithms are provided in Crisan and Doucet (2000), Beskos et al. (2016), and Giraud et al. (2017).

2.4.2 Stopping criterion

We present a stopping rule that controls the number of Metropolis-Hastings updates within the mutation step. This provides an automatic heuristic determining when to stop the mutation stage, and it can also eliminate unnecessary computer model runs. The stopping rule is based on the Bhattacharyya distance (Bhat-

tacharyya, 1946), $D_B(p, q)$, which measures the similarity between two distributions, $p(\theta)$ and $q(\theta)$. We first evaluate the stopping criterion after $2k$ Metropolis-Hastings updates; if the criterion is not met, then we re-evaluate after k subsequent updates.

Consider $\theta_t^{i,k}$, the ith particle, or parameter setting, after the kth mutation step of the Metropolis-Hastings algorithm during cycle number t. Let $\boldsymbol{\theta_t^k} = \{\theta_t^{1,k}, ..., \theta_t^{n,k}\}$ denote the set of parameters $\theta_t^{i,k}$'s. Let $h(\theta_t^{i,k})$ be the target metric of interest evaluated at parameter setting $\theta_t^{i,k}$, in this case, the Antarctic ice sheet contribution to sea level change in 2100. Let $\mathbf{h}(\boldsymbol{\theta_t^k}) = \{h(\theta_t^{1,k}), ..., h(\theta_t^{n,k})\}$ denote the set of target metrics $h(\theta_t^{i,k})$'s.

At mutation update $2k$, we partition the range spanned by two sets of target metric samples – $\mathbf{h}(\boldsymbol{\theta_t^k})$ and $\mathbf{h}(\boldsymbol{\theta_t^{2k}})$ – into m non-overlapping blocks of equal width. Then, we compute the real-valued Bhattacharyya distance $D_B(\mathbf{h}(\boldsymbol{\theta_t^k}), \mathbf{h}(\boldsymbol{\theta_t^{2k}})) = -\ln\left(\sum_{i=1}^{n}\sqrt{p_i q_i}\right)$ where p_i and q_i are the proportion of samples, from $\mathbf{h}(\boldsymbol{\theta_t^k})$ and $\mathbf{h}(\boldsymbol{\theta_t^{2k}})$ respectively, that lie within the ith partition. The mutation step proceeds until $D_B(h(\theta_k), h(\theta_{2k})) < \epsilon_B$, the stopping criterion. If the stopping criterion is not fulfilled, we run k additional Metropolis-Hastings updates and evaluate the stopping criterion at iterations $3k$ and $2k$. We repeat this until the stopping criterion is met. We obtain the threshold ϵ_{BD} through a Monte Carlo simulation run prior to the calibration study. Section 2.4.4 discusses tuning for k, ϵ_B, and m.

2.4.3 Adaptive incorporation schedule

In Equation 2.2, we introduce a standard incorporation schedule $\gamma = (\gamma_0, ..., \gamma_T)$. In the standard implementation, the user must select the total number of sampling-importance-resampling cycles (T) and the likelihood incorporation increments γ_t for $t = (0, ..., T)$. Past research proposed novel methods to adaptive choose the incorporation schedule, γ_t, yet maintain a constant number of cycles, T (Nguyen, 2014; Kalyanaraman et al., 2016). Here, we introduce an adaptive incorporation schedule that automatically determines both the total number of sampling-importance-resampling cycles, T, and incorporation schedule,γ. Introducing the adaptive incorporation schedule into the particle-based calibration framework provides computational and practical benefits by (1) reducing the number of computer

model evaluations; (2) decreasing the overall calibration wall times; and (3) simplifying implementation for the user.

The adaptive incorporation schedule proceeds as follows. On initialization, we set the initial incorporation increment γ_0 to 0. We draw the initial set of particles θ_0 from $\pi_0(\theta|Z) \propto L(\theta|Z)^0 p(\theta) = p(\theta)$, the prior distribution of model parameters. For cycle $t = 1, 2, 3, ...,$ we calculate the full likelihood $L(\theta_{t-1}^{(i)}|Z)$ for $i = 1, ..., N$ where $\theta_{t-1}^{(i)}$ denotes the parameter samples from the previous cycle $t-1$. For computational efficiency, we reuse the likelihood evaluations from the previous cycle. Next, we find the optimal γ_t that returns an effective sample size (ESS) of ESS_{thresh} or a sample size closest to ESS_{thresh}:

$\gamma_t = \text{argmin}_\gamma\{(ESS_\gamma - ESS_{thresh})^2\}$

, where $\gamma \in (\gamma_{min}, 1 - \gamma_{t-1})$, γ_{min} is a previously set minimum incorporation value, $ESS_\gamma = \sum_{i=1}^{N} \frac{1}{w_t^{(i)2}}$, and $w_t^{(i)} \propto L(\theta_t^{(i)}|Z)^\gamma$. Note that we can lower computational costs by evaluating the full likelihood $L(\theta_0^{(i)}|Z)$ only once before the optimization.

We stop the scheduling algorithm when $\sum_{i=1}^{t} \gamma_t = 1$. This occurs when the entire likelihood has been incorporated, and the target distribution has evolved to the full posterior distribution $\pi(\theta|Z)$. Note at each cycle t, we set the incorporation increment (γ_t) to be between γ_{min} and $1 - \sum_{i=1}^{t} \gamma_t$. In Section 2.4.4, we describe how to set the minimum incorporation increment γ_{min} and the threshold effective sample size, ESS_{thresh}.

Adaptive likelihood incorporation schedule

1. Initialization: At $t = 0$, set $\gamma_0 = 0$.
2. When $t > 0$ and $\sum_{i=1}^{t-1} \gamma_i < 1$
   - Compute $L(\theta_{t-1}^{(i)}|Z)$ for $i = 1, ..., N$
   - Set $\gamma_t = \text{argmin}_\gamma\{(ESS_\gamma - ESS_{thresh})^2\}$, where $ESS_\gamma = \sum_{i=1}^{N} \frac{1}{w_t^{(i)2}}$, $w_t^{(i)} \propto L(\theta_t^{(i)}|Z)^\gamma$, and $\gamma \in (\gamma_{min}, 1 - \gamma_{t-1})$.
   - γ_{min} is a predetermined minimum incorporation value
3. When $t > 0$ and $\sum_{i=1}^{t-1} \gamma_i = 1$: Stop Calibration

Algorithm 1: Fast Particle-based Calibration

Data: Z

Initialization:

Draw $\theta_0^{(i)} \sim p(\theta)$ for particles $i = 1, ..., N$.

Set $w_0^{(i)} = 1/N$, $\gamma_0 = 0$, and K;

for *cycles* $t = 1, ..., T$ **do**

1. Compute full likelihood:

Calculate $L(\theta_{t-1}^{(i)}|Z)$ for $i = 1, ..., N$;

2. Select optimal likelihood incorporation increment γ_t:

Set $\gamma_t = \text{argmin}_\gamma\{(ESS_{\gamma_t} - ESS_{thresh})^2\}$, where $\gamma \in (0.1, 1 - \sum_{i=1}^{t-1} \gamma_{t-1})$

Note: $ESS_{\gamma_t} = \sum_{i=1}^{N} \frac{1}{w_t^{(i)2}}$ and $w_t^{(i)} \propto L(\theta_t^{(i)}|Z)^{\gamma_t}$;

3. Compute importance weights:

$w_t^{(i)} \propto w_{t-1}^{(i)} \times L(\theta_t^{(i)}|Z)^{\gamma_t}$;

4. Re-sample particles:

Draw $\theta_t^{(i)}$ from $\{\theta_{t-1}^{(1)}, ..., \theta_{t-1}^{(N)}\}$ with probabilities $\propto \{w_t^{(1)}, ..., w_t^{(N)}\}$;

5. Set intermediate posterior distribution:

Set $\pi_t(\theta|Z) \propto L(\theta_i|Z)^{\tilde{\gamma}}\pi(\theta)$, where $\tilde{\gamma} = \sum_{j=1}^{t} \gamma_j$;

6. Mutation:

Using each particle $(\theta_t^{(1)}, ..., \theta_t^{(N)})$ as the initial value, run N chains of an MCMC algorithm with target distribution $\pi_t(\theta|Z)$ for $2K$ iterations

7. Check stopping criterion:

Compute $\delta_B = D_B(h(\theta_t^K), h(\theta_t^{2K}))$;

if $\delta_B < \epsilon_B$ **then**

Set $\theta_t^{(i)} = \theta_t^{(i),2K}$;

else

Run K additional updates and re-evaluate stopping criterion

Continue until stopping criterion is met

8. Stop when full likelihood is incorporated;

if $\sum_{i=1}^{N} \gamma_t = 1$ **then**

End Algorithm;

else

Reset weights: $w_t^{(i)} = 1/N$ for particles $i = 1, ..., N$;

Set t=t+1 and return to Step 1;

2.4.4 Tuning the algorithm

Much of the algorithm above is automated. However, the user needs to choose: (1) the total number of particles, N; (2) the number of Metropolis-Hastings updates run before checking the stopping criterion, K; (3) the minimum incorporation γ_{min}; and (4) the effective sample size threshold ESS_{thresh}. (1) and (2) should be set based on the amount of available computational resources, but our simulation study results favor having more particles N than longer Metropolis-Hastings updates K. We chose 2015 particles, which requires 56 nodes with 36 processors per node; thereby leaving one processor to execute master tasks. We set the reference length k for the Metropolis-Hastings updates to be 7. Based on simulation experiments, the empirical distribution of particles stabilize after 10 to 15 updates. In this study, we set the floor for the incorporation increment, γ_{min} to be 0.1 so that at each cycle, the weights for the importance sampling step is at least $L(\theta|Z)^{0.1}$.

The automated likelihood tempering schedule (Section 2.4.3) ensures that the effective sample size (ESS) of the final particles does not fall below a pre-determined threshold ESS_{thresh}. For moderate-dimensional parameter spaces (5-20), the effective sample size is important as it is an indicator of the discrepancy between the true target distribution and the particle-based empirical distribution (cf. Doucet et al., 2001; Gordon et al., 1993). A low ESS suggests that only a few particles have any significant weight, and it is often indicative of weight degeneracy and a poor approximation of the target distribution (Kong, 1992). A suitable ESS can be obtained by minimizing ρ, the second moment of the Radon-Nikodym derivative between the target and the proposal distribution (Whiteley et al., 2016; Kong, 1992), generating more sophisticated proposal distributions via implicit sampling (Morzfeld et al., 2015), and examining distances between target and proposal distributions within an intrinsic dimension (Agapiou et al., 2017). Other research (Martino et al., 2017) point to alternative definitions of the ESS than the traditional method based on the variance of the weights (Liu and Chen, 1998). In this study, we utilize the common definition of ESS (Kong, 1992), which is based on the variance of the importance weights, and we set $ESS_{thresh} = \frac{N}{2}$, which is the typical threshold used by many sequential Monte Carlo methods (Del Moral et al., 2006) prior to resampling.

We obtain ϵ_{BD} as follows. Prior to running the calibration algorithm, we ob-

tained samples of a target metric (Antarctic Ice Sheet contribution to sea level rise in 2100) from an initial survey of computer model runs. Let μ and σ^2 denote the sample mean and variance of the target metric mentioned above. We generate a collection of B samples of size n, denoted as $\mathbf{x} = \{x_1, ..., x_B\}$. Here, $x_b \sim \mathcal{N}(\mu, \sigma^2)$, with μ and σ^2 previously defined. Let $x_{base} \sim \mathcal{N}(\mu, \sigma^2)$ be a baseline sample for calculating the Bhattacharrya distance. We calculate $D_B(x_b, x_{base})$ for $b = 1, ..., B$, and set ϵ_{BD} to be the 0.975 quantile. In this study, we chose $B = 1000$ and the number of partitions $m = 200$.

We calibrate the PSU3D-ICE model using Cheyenne (Computational and Information Systems Laboratory, 2017), a 5.34-petaflops high performance computer operated by the National Center for Atmospheric Research (NCAR). Parallelized operations, such as calculating importance weights and mutation, proceed via message passing interface (MPI). To limit communication costs, we build the ice sheet model and load the relevant datasets separately on each processor.

2.4.5 Computational advantages and limitations

We take advantage of the embarrassingly parallel nature of the importance sampling and mutation steps to reduce wall time. In our approach, the Metropolis-Hastings updates in the mutation stage are the primary drivers of computational cost. To address this cost, we propose an automated stopping rule for the mutation stage. We also introduce an adaptive likelihood incorporation schedule that automatically selects an efficient number of sampling-importance-resampling cycles. The stopping rule and adaptive likelihood incorporation schedule simplifies implementation for the user (due to automation) and reduces the number of computer model runs needed for calibration.

Our approach is a viable alternative to existing calibration methods, which may be computationally infeasible. MCMC-based calibration methods using the computer model is computationally prohibitive due to the sequential nature of MCMC algorithms. Emulation-calibration methods, while efficient for expensive computer models, do not easily scale to problems with many parameters (say more than five or six for this model). Also, multiple-try MCMC methods (Liu et al., 2000), a mixture of importance sampling and MCMC, may incur large

costs because several parallel processes must be initialized and terminated at each iteration of the MCMC chain. Multiple-try MCMC may experience slow mixing, especially when the Markov chain moves to the low-probability regions of the target distribution distribution (Martino, 2018).

While our method has many computationally advantages, we note that the heavy parallelization in our approach requires access to and the ability to work with high performance computing resources. Given our current computing resources, our method is ideally suited to models that run between six seconds and 15 minutes. For models with longer run times, the computational costs remain prohibitive. MCMC algorithms may be feasible and simpler to implement for models with shorter run times. As is the case with parallel computing methods, communication costs must be small relative to the computer model run times; otherwise we would not reap the benefits of our approach.

2.5 Simulated example and results

In this section, we calibrate a simple computer model using three different methods. We simulate a data set of size $n = 300$ where the spatial locations s_i for $i = 1, ..., n$ are in the unit domain $[0, 1]^2$. We generate the data via a modified version of the computer model presented in Bayarri et al. (2007). We construct a simple computer model as follows:

$$Y(s_i, \theta) = 5 \times \exp\{-\theta(\text{lat}_i \times \text{lon}_i)\},$$

where $Y(s_i, \theta)$ is a real-valued computer model output at model parameter setting θ and at a spatial location specified by lat_i and lon_i, which represent the latitude and longitude of the ith location. The true process includes a data-model discrepancy term $\delta(s_i)$, which is defined as $\delta(s_i) = -1.5 \times (\text{lat}_i \times \text{lon}_i)$, and iid observational error $\epsilon_i \sim \mathcal{N}(0, \sigma^2_\epsilon)$. For this example, we set $\theta = 1.7$ and $\sigma^2_\epsilon = 0.5$. To generate the observational data, $Z(s_i)$, we combine the computer model output $Y(s_i, \theta)$, the data-model discrepancy, $\delta(s_i)$, and the observational error, ϵ_i, as follows:

$$Z(s_i) = Y(s_i, \theta) + \delta(s_i) + \epsilon_i.$$

We model the observations as

$$Z(s_i) = 5 \times \exp\{-\theta(\text{lat}_i \times \text{lon}_i)\} + \delta(s_i) + \epsilon_i,$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ are the iid observational errors. Since the actual form of the discrepancy term is unknown, we model the discrepancy $\delta(s_i)$, as a zero-mean Gaussian process, $\delta(s) \sim \mathcal{GP}(0, \Sigma_\delta(\xi_\delta))$, where ξ_δ is a vector containing the covariance parameters. To allow for some roughness of the process between spatial locations we choose an exponential covariance function $\Sigma_\delta(\xi_\delta) = \sigma_\delta^2 \exp\left(-\frac{|s_i - s_j|}{\phi_\delta}\right)$ with $\xi_\delta = (\phi_\delta, \sigma_\delta^2)$. To complete the Bayesian framework, we use the prior distributions: $\theta \sim \mathcal{N}(0, 100)$, $\sigma_\epsilon^2 \sim \mathcal{IG}(2, 2)$, $\phi_\delta \sim \mathcal{U}(0.01, 1.5)$, and $\sigma_\delta^2 \sim \mathcal{IG}(2, 2)$.

We compare results from three calibration methods: (1) MCMC-based, (2) standard particle-based, and (3) adaptive particle-based. In the MCMC-based method, we generated $100,000$ samples from $\pi(\theta, \phi_\delta, \sigma_\delta^2, \sigma_\epsilon^2 | Z)$ via the Metropolis-Hastings algorithm. Next, the standard and adaptive particle-based calibration methods use $N = 2000$ particles to approximate $\pi(\theta, \phi_\delta, \sigma_\delta^2, \sigma_\epsilon^2 | Z)$. For the standard particle-based method, we set the total number of cycles to be 10, and establish a uniform likelihood incorporation $\gamma = (\gamma_1, ..., \gamma_{10})$, where $\gamma_t = 0.1$ for $t = 1, ..., 10$. We run $K = 100$ Metropolis-Hastings updates for each mutation cycle. In the adaptive particle-based calibration approach, our algorithm automatically chose four cycles with a likelihood incorporation schedule $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (0.100, 0.148, 0.2743, 0.4777)$ using the adaptive likelihood incorporation schedule. For each mutation step, we run batches of $K = 5$ Metropolis-Hastings updates until the stopping criterion is met.

All three methods yield comparable calibration results (see Table 2.1); however, our adaptive particle-based approach exhibits a considerable speedup in computation. For the model parameter, θ, calibration via MCMC (the "gold standard") provides estimate $\hat{\theta}_{mcmc} = 2.04$ and 95% credible interval bounds $(1.06, 3.17)$. Similarly, the standard particle-based approach generates estimate $\hat{\theta}_{std} = 2.04$ with 95% credible interval bounds $(1.03, 3.11)$ and the adaptive particle-based approach yields estimate $\hat{\theta}_{adapt} = 2.04$ with 95% credible interval bounds $(1.05, 3.14)$. The adaptive particle-based approach has considerably shorter wall times due to fewer computer model evaluations. To illustrate, the adaptive approach requires just $10 \times 4 = 40$ sequential computer model runs, as opposed to $10 \times 200 = 2000$

runs for the standard particle-based approach and $100,000$ for the MCMC-based approach.

Table 2.1. Simulated example calibration results for three calibration methods: (1) Adaptive particle-based; (2) Standard particle-based; and (3) MCMC with full model. All three approaches yield comparative results.

	θ	ϕ_δ	σ^2_δ	σ^2_ϵ
Adaptive Particle (Est)	2.04	1.22	0.78	0.44
Adaptive Particle (95% CI)	(1.05,3.14)	(0.83,1.50)	(0.36,1.32)	(0.36,0.52)
Standard Particle (Est)	2.04	1.22	0.80	0.44
Standard Particle (95% CI)	(1.03,3.11)	(0.81,1.50)	(0.32,1.33)	(0.35,0.51)
MCMC-Based (Est)	2.04	1.21	0.79	0.44
MCMC-Based (95% CI)	(1.06,3.17)	(0.80,1.50)	(0.34,1.33)	(0.36,0.52)

2.6 Application to the PSU3D-ICE model

Here we provide specifics for calibrating the PSU3D-ICE model and discuss how our method provides key computational benefits over existing calibration approaches. We also summarize results from a comparative analysis of three calibration methods within the context of the PSU3D-ICE model. The efficiency of our computational approach allows us to study the effect of observations from the Pliocene on parameter calibration and projections of sea level rise and also enables us to conduct a prior sensitivity analysis.

2.6.1 Calibrating PSU3D-ICE

We calibrate 11 model parameters using both paleoclimate records and modern observations from satellite imagery (Section 2.2). For the paleoclimate records, modern volume, and modern grounded ice area, we use independent truncated normal distributions. The upper and lower ranges for the truncated normal likelihood functions are based on domain area expertise and past studies (Section 2.2.2).

We calibrate the PSU3D-ICE model using five observations: (1) Z_{plio}, the Antarctic ice sheet's contribution to sea level change (m) in the Pliocene; (2) Z_{lig}, contribution in the Last Interglacial Age (m); (3) Z_{lgm}, contribution in the

Last Glacial Maximum (m); (4) Z_{vol}, the Antarctic ice sheet's total ice volume in the modern era (km^3) ; and (5) Z_{area}, total grounded ice area in the modern era (km^2). We also use observations of ice occurrence taken at 10 strategic point in the Antarctic Ice Sheet. Here, $Z_{spat} = (Z_{spat,1}, ..., Z_{spat,10})$. All ten locations have ice presence; so, $Z_{spat,i} = 1$ for locations $i = 1, ..., 10$.

Likelihood

For the observational dataset,
$\mathbf{Z} = (Z_{plio}, Z_{lig}, Z_{lgm}, Z_{vol}, Z_{area}, Z_{spat,1}, ..., Z_{spat,10})$, we define a likelihood function using truncated normal distributions and indicator functions. For the modern volume, modern total grounded area, and paleoclimate records, we use independent truncated normal distributions as the observational model. $TN(\mu, \sigma^2, \alpha, \beta)$ denotes a truncated normal distributions with the mean (μ), variance (σ^2), upper bound (α), and lower bound (β).

$$Z_{plio} \sim \mathcal{TN}(\mu = Y(\theta)_{plio}, \sigma^2 = 30^2, \alpha = Y(\theta)_{plio} - 10, \beta = Y(\theta)_{plio} + 10)$$

$$Z_{lig} \sim \mathcal{TN}(\mu = Y(\theta)_{lig}, \sigma^2 = 10^2, \alpha = Y(\theta)_{lig} - 2, \beta = Y(\theta)_{lig} + 2)$$

$$Z_{lgm} \sim \mathcal{TN}(\mu = Y(\theta)_{lgm}, \sigma^2 = 20^2, \alpha = Y(\theta)_{lgm} - 5, \beta = Y(\theta)_{lig} + 5)$$

$$Z_{vol} \sim \mathcal{TN}(\mu = Y(\theta)_{vol}, \sigma^2 = 1.6\times10^{15}, \alpha = Y(\theta)_{vol} - 2.5\times10^{15}, \beta = Y(\theta)_{vol} + 2.5\times10^{15})$$

$$Z_{area} \sim \mathcal{TN}(\mu = Y(\theta)_{ar}, \sigma^2 = 0.6\times10^{12}, \alpha = Y(\theta)_{ar} - 1.5\times10^{12}, \beta = Y(\theta)_{ar} + 1.5\times10^{12})$$

The second set of observations are binary occurrences of ice at 10 strategically placed points on the Antarctic ice sheet (Lee et al., 2019). For these observations, we use indicator functions as the observational model as follows:

$$Z_{spat} \sim \prod_{i=1}^{10} \mathbb{I}(Y(\theta)_{spat,i} = Z_{spat,i}),$$

where $Y(\theta)_{spat,i}$ denotes the model spatial output for a model run using parameters θ.

Priors

We set the prior distributions for the 11 model parameters based on expert knowledge. Five model parameters - CALVNICK, TAUASTH, CALVLIQ, CLIFFVMAX, FACEMELTRATE - have uniform prior distributions. Here, $\theta \sim U(\alpha, \beta)$, where α and β denote the upper and lower bounds of the uniform distribution. The prior distributions are as follows:

- $\theta_{CALVNICK} \sim \mathcal{U}(0, 2)$
- $\theta_{CLIFFVMAX} \sim \mathcal{U}(0, 12000)$
- $\theta_{TAUASTH} \sim \mathcal{U}(1000, 5000)$
- $\theta_{CALVLIQ} \sim \mathcal{U}(0, 200)$
- $\theta_{FACEMELTRATE} \sim \mathcal{U}(0, 20)$

Six parameters - OCFACMULT, OCFACMULTASE, CRHSHELF, ENHANCESHEET, ENHANCESHELF, CRHFAC - have log-uniform prior distributions. Here, $\theta \sim LU(base, \alpha, \beta)$, which implies $log_{base}(\theta) \sim U(\alpha, \beta)$ where α and β denote the upper and lower bounds of the uniform distribution. The prior distributions are as follows:

- $\log_{10}(\theta_{OCFACMULT}) \sim \mathcal{U}(-0.5, 0.5)$
- $\log_{10}(\theta_{OCFACMULTASE}) \sim \mathcal{U}(0, 1)$
- $\log_{10}(\theta_{CRHSHELF}) \sim \mathcal{U}(-7, -4)$
- $\log_{10}(\theta_{ENHANCESHEET}) \sim \mathcal{U}(-1, 1)$
- $\log_{0.3}(\theta_{ENHANCESHELF}) \sim \mathcal{U}(-1, 1)$
- $\log_{10}(\theta_{CRHFAC}) \sim \mathcal{U}(-2, 2)$

We can estimate the data-model discrepancy as an additive model bias, $\alpha \in \mathbb{R}$, such that our observational model (2.1) is modified to be $Z = Y(\theta) + \alpha + \epsilon$. For observations that are discontinuous in time, past ice sheet calibration studies (Edwards et al., 2019; Williamson et al., 2013; Ruckert et al., 2017) model the discrepancy term as a tolerance to the observation measurement error, which follows the zero-mean Gaussian process framework provided in Kennedy and O'Hagan (2001). For the PSU3D-ICE model, we find that calibration with and without the discrepancy term yields very similar results.

2.6.2 Computational benefits of our approach

Our adaptive particle-based approach greatly reduces calibration wall times compared to using an all-at-once random-walk Metropolis-Hastings algorithm as in past ice-sheet calibration studies (cf. Ruckert et al., 2017; Bakker et al., 2016; Petra et al., 2014). Our fast calibration approach had a total wall-time of $\sim$ 6.5 hours and evolved 2015 particles for an effective sample size (ESS) of 1533. For the MCMC-based calibration approach, it would be computationally prohibitive to generate a large enough sample with a similar ESS. Instead, we estimate the time to generate an ESS of 1533. We ran the Metropolis-Hastings algorithm for 12 days to generate 1500 samples. We calculated the effective sample size per hour (Jones et al., 2006) for each model parameter and then project the time required to obtain an ESS of 1533, the ESS from the particle-based approach. It would require 12 to 18 months running the Metropolis-Hastings algorithm to generate the same ESS as our particle-based approach.

The computing times are based on the PSU3D-ICE model run at 80 km spatial resolution and an adaptive temporal resolution with a baseline timestep of 8 years. Run times are for the NCAR Cheyenne HPC system with 2.3-GHz Intel Xeon E5-2697V4 Broadwell processors. Note that in practice, computation times for the particle-based methods can be slightly higher due to initialization and communications costs inherent to parallelized computing. Reduction of initialization and communication costs is an active area of research with novel methods in development (Ballard et al., 2016; Fan et al., 2018). Note that the computation times for the MCMC-based approach are quite optimistic as we initialized the Markov chain and set the proposal distribution using all samples generated from our particle-based approach. In general, MCMC algorithms would not have access to these particles, and would therefore likely require even more iterations of the MCMC algorithm to achieve the desired ESS.

2.6.3 Comparisons to other calibration approaches

We conduct a comparative study between our particle-based calibration approach and competing emulation-calibration methods (see Supplement (Lee et al., 2019) for details). We calibrate the PSU3D-ICE model using three methods:

1. A low-dimensional emulation-calibration approach: This approach varies only three parameters – OCFACMULT, CALVLIQ and CLIFFVMAX – and fixes the remaining eight parameters at scientifically justified values provided by our expert on ice sheets (DP). We include this approach because reducing the number of parameters is a common way to address computational challenges associated with calibration with long model run times (e.g. Edwards et al., 2019; Chang et al., 2014; Sacks et al., 1989). We chose these three parameters because they are considered to be important in modeling the long-term evolution of the Antarctic ice sheet (Edwards et al., 2019; DeConto and Pollard, 2016). We train a Gaussian process emulator using 512 design points and use the squared exponential covariance function to represent the dependence between the design points. For the experimental design, we use a full factorial design with eight equally spaced points for each model parameter.

2. A high-dimensional emulation-calibration approach: We calibrate all 11 selected parameters of the PSU3D-ICE model. We train a Gaussian process emulator using 512 design points generated via Latin Hypercube Design (LHC). Similar to the low-dimensional case, we use an exponential covariance function to model the dependence between design points. Emulation and calibration details are provided in the Supplement (Lee et al., 2019).

3. Our particle-based approach: We use our heavily parallelized particle-based approach to calibrate all 11 selected parameters.

For the first method, we find that by fixing eight of eleven parameters, we greatly constrain the parameter space and thereby underestimate the parametric uncertainty underlying the ice sheet model. Projections for the Antarctic sea level contribution in 2100-2500 are much lower and overconfident compared to those from our particle-based approach (Figure 2.3). For the second method, the limited amount of design points (training data) generates an inaccurate surrogate model as shown by the large out-of-sample cross-validated mean squared prediction error (Supplement (Lee et al., 2019)). This calls into question the parameter estimates as well as the resulting projections. As shown in Figure 2.2, the second approach produces extremely sharp posterior distributions for two key model parameters,

CLIFFVMAX and TAUASTH, which is inconsistent with the parameter estimates from the particle-based approach.

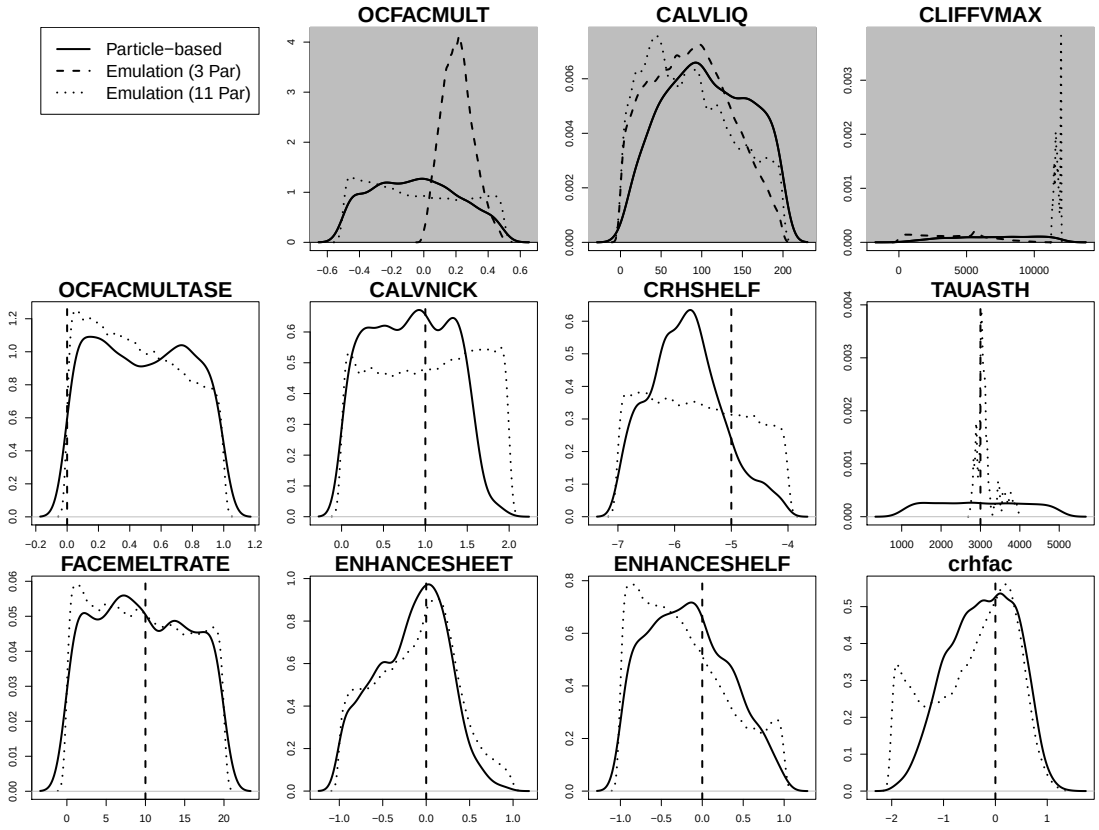

Figure 2.2. Posterior densities of model parameters using the adaptive particle-based approach (solid line), emulation calibration with three parameters (dashed line), and emulation calibration with 11 parameters (dotted line). Three-parameter emulation-calibration experiment use model parameters OCFACMULT, CALVLIQ, and CLIFFVMAX. The 11-parameter emulation-calibration experiment include all model parameters. Shaded panels denote parameters used in the three-parameter emulation-calibration experiment.

Figure 2.4 compares the posterior densities of projections and hindcasts for the three-parameter emulation-calibration approach and our 11-parameter particle-based method. Note that the three-parameter emulation-calibration approach (striped blue shading) underestimates the tail-area risk, or the 99-th% quantile, for sea level projections compared to our approach (striped red shading). By calibrating more parameters, we can expect the tail-area risk to increase by a factor of 74 in 2100 and 65 in 2300.

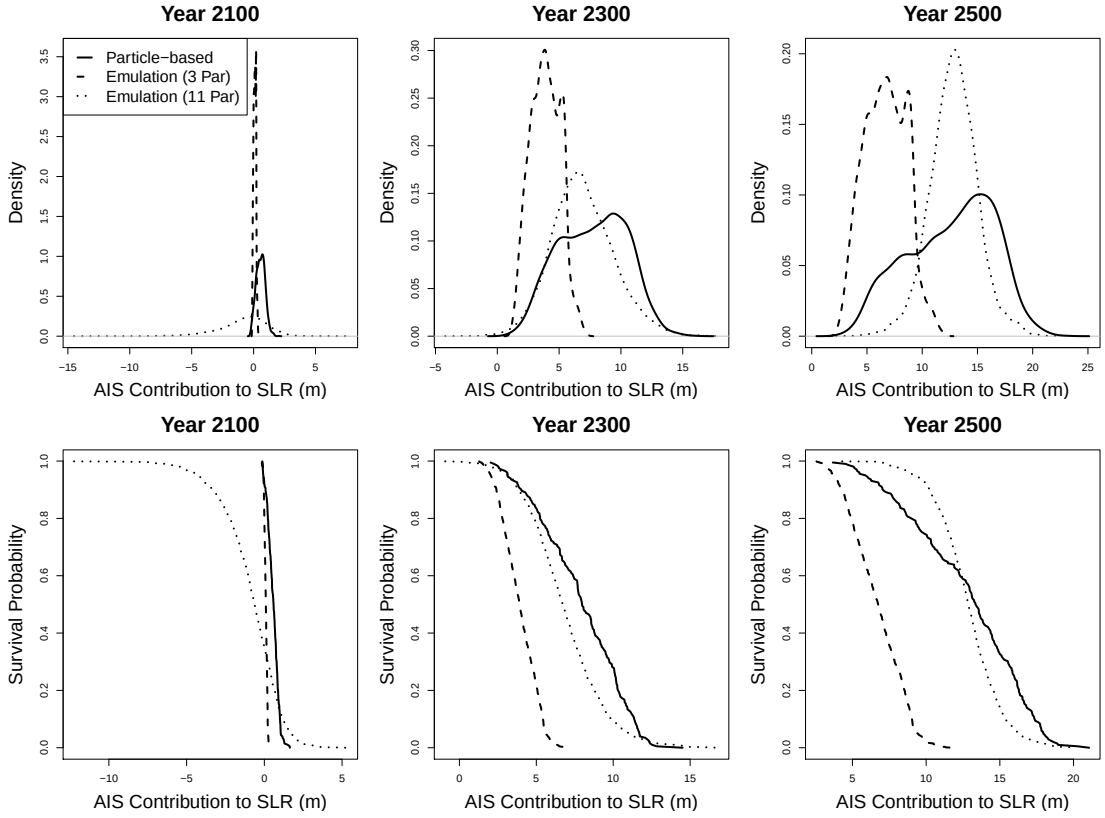

Figure 2.3. (Top Panel) Posterior densities of the projected Antarctic ice sheet's contribution to sea level change in 2100, 2300, and 2500 using the adaptive particle-based approach (solid line), emulation calibration with three parameters (dashed line), and emulation calibration with 11 parameters (dotted line). (Bottom Panel) Empirical survival functions of the projected Antarctic ice sheet's contribution to sea level change in 2100, 2300, and 2500 using the adaptive particle-based approach (solid line), emulation calibration with three parameters (dashed line), and emulation calibration with 11 parameters (dotted line). Three-parameter emulation results in sharper densities centered on distinctively lower point estimates. The 11-parameter emulation-calibration approach results in highly uncertain projections.

The three-parameter emulation calibration required 1.5 minutes to fit the Gaussian process emulator using 12 processors on the Cheyenne HPC system and $\sim$1.5 hours to generate 500k samples via MCMC from the posterior distribution. The 11-parameter emulation calibration required 10 minutes to fit the emulator using 12 processors on the Cheyenne HPC system and $\sim$1.5 hours to generate 500k samples via MCMC from the posterior distribution.

2.6.4 The effect of deep time observations on projections

Calibration can be improved by considering an important source of uncertainty, the state of the Antarctic ice sheet during the Pliocene era (Dolan et al., 2018; Salzmann et al., 2013; Dutton et al., 2015). There is some evidence that the Antarctic ice sheet experienced fluctuations in volume during the Pliocene era (Naish et al., 2009). Other studies suggest that at peak warming episodes during the Pliocene era, the Antarctic ice sheet had a lower volume, contributing to higher sea level rise (Cook et al., 2014; Dolan et al., 2011; Dowsett and Cronin, 1990; Pollard and DeConto, 2009; Pollard et al., 2015; De Boer et al., 2014). However, the maximum Antarctic ice retreat and sea level rise contribution during the Pliocene remains largely uncertain (Dutton et al., 2015; Rovere et al., 2014).

We examine whether the width of the Pliocene observation windows (5 m to 25 m, 5 m to 10 m, 10 m to 25 m) has an influence on sea level projections and parameter estimation. (See Supplement (Lee et al., 2019) for details on how these windows affect the likelihood function.) Our results demonstrate that information regarding the nature of the Antarctic ice sheet during the Pliocene era has a strong influence on sea level projections. Figure 2.5 illustrates how the posterior densities for two key model parameters (CALVLIQ and CLIFFVMAX) differ under the three Pliocene windows. Both parameters influence ice dynamics inherent to marine cliff instability (MICI) – hydrofracturing due to surface melt (CALVLIQ) and structural failure of tall ice cliffs (CLIFFVMAX). As shown in Figure 2.6, increasing the Pliocene window from the range 5 m to 10 m to the range 10 m to 25 m requires more aggressive MICI (larger values of these parameters); hence resulting in higher projections of sea level rise (e.g. exceeding 3 m in 2300). If we are very uncertain about the Pliocene (represented by a very large window of 5 m to 25 m), the resulting sea level projections in 2300 also become highly uncertain (95% credible interval of 1.2 m to 12.4 m), compared to projections from narrower windows of 5 m to 10 m (95% credible interval of 1.2 m to 11.5 m) or 10 m to 25 m (95% credible interval of 3.0 m to 12.9 m). The experiments using low (5 m to 10 m) and high (10 m to 25 m) Pliocene windows utilized subsets of the samples generated from the main calibration, and the corresponding sub-samples had an effective sample size (ESS) of 891 and 642, respectively.

2.6.5 Sensitivity to model parameter priors

Calibration results may exhibit sensitivity to the choice of the model parameters' prior distributions (cf. Jackson et al., 2015; Reese et al., 2004), especially for sparse observational records. This constitutes an important source of second-order, or deep uncertainty, an important factor in the design of risk management strategies (Keller and McInerney, 2008). To examine prior sensitivity, we calibrate the ice sheet model using two sets of prior distributions which are in the form of uniform or log-uniform distributions. One set of priors has a much wider range (large difference between upper and lower bounds) than the other. The much wider ranges represent physically possible parameter values that do not violate any fundamental physical laws, and the narrower ranges represent values that yield reasonable model behavior found in many years of unstructured tuning by the model developers (Pollard and DeConto, 2012a). We provide additional details in the Supplement (Lee et al., 2019).

The choice of prior distributions has a notable effect on parameter estimates (Figure 2.7) and sea level projections (Figure 2.8 and Table 2.2). Note that constraining the model parameters *a priori* may underestimate sea level projections. However, overly wide prior distributions may permit physically unrealistic outcomes. Hence, it is important to carefully construct prior distributions based on domain area expertise, as we have in this manuscript. In particular, changing the prior on the parameter CLIFFVMAX – wastage rate for unstable marine ice cliffs – can have a strong impact on projections. For a prior range of 0 km/year to 12 km/year, the 95% credible interval for the Antarctic ice sheet's contribution to sea level rise in 2300 is 1.2 m to 12.4 m. A wider prior range of 0 km/year- to 600 km/year results in considerably higher projection uncertainty denoted by a 95% credible interval of 0.7 m to 21.0 m. For the experiment using the wide priors, our particle-based calibration approach utilized 2015 particles to obtain an effective sample size (ESS) of 1583.

Table 2.2. Antarctic ice sheet's projected contribution to sea level change in 2100-2500 after calibration using narrow and wide prior distributions.

Prior	Year 2100	Year 2200	Year 2300	Year 2400
Narrow	0.4 (-0.3, 1)	3.8 (0.1, 6.7)	7.9 (1.2, 12.4)	10.6 (2.5, 15.5)
Wide	1.8 (-0.4, 5.5)	10 (-0.2, 19.5)	13.9 (0.7, 21)	15.5 (1.8, 21.8)

The wider range for CLIFFVMAX explores a fundamental uncertainty in MICI – the rate at which very tall ice cliffs will disintegrate back into the ice sheet interior. If grounding lines retreat into the interior of deep Antarctic basins, the exposed ice cliffs will be taller than any observed today, and the wastage velocities (CLIFFVMAX) could conceivably be much greater than the approximately 12 km per year observed today at the ice fronts of major Greenland glaciers (which might not even be approximate analogs for MICI, being driven instead mainly by buoyant calving; Murray et al. (2015)). The bimodal character of the posterior densities in the top panels of Figure 2.8 for 2300 and 2500 are due to the very large CLIFFVMAX range. The upper peak centered on around 20 m is produced by CLIFFVMAX values of approximately 100 km per year and above, which produce collapse of almost all marine ice in both East and West Antarctica. The lower peak centered on around 5 m occurs for many lower CLIFFVMAX values, for which the more vulnerable West Antarctic ice sheet collapses, but marine basins in East Antarctica do not retreat.

2.7 Discussion

2.7.1 Summary

We present a novel particle-based approach to calibrate the 80 km resolution PSU3D-ICE model. We show that our approach provides good approximations and drastically reduces overall calibration wall times by heavily parallelizing the sequential Monte Carlo algorithm, and carefully tuning the algorithm to drastically reduce the number of sequential model evaluations. Our algorithm is applicable to a broad class of models that have a moderate run time (given our computing resources, between a few seconds and several minutes) and a moderate number of model parameters (in our case between 5 and 20).

We use this new method to assess the impacts of neglecting parametric uncertainties on sea level projections. Emulation-calibration methods using fewer parameters yield lower and more overconfident projections of sea level rise than using more parameters through the particle-based calibration approach. This method includes the recent study of Edwards et al. (2019), who found that the important

mechanism of marine ice cliff instability (MICI) is not necessary to capture past variations. In this case, future sea level projections are considerably lower. In contrast, our new approach that accounts for more parametric uncertainties suggests that MICI may still be important and future sea level projections may be much higher, especially considering potential Pliocene windows. Using emulation-calibration in a high-dimensional parameter space induces considerable emulator-model discrepancy and can result in large projection uncertainties. Our method utilizes the actual ice sheet model; thereby preserving the highly non-linear ice dynamics as well as the complex interactions between model parameters. This has clear policy-relevant implications because projections from ice sheet models inform economic and engineering assessments (cf. Sriver et al., 2018; Diaz and Keller, 2016; Johnson et al., 2013).

Our approach enables calibration experiments that were computationally prohibitive using current calibration methods. First, assuming different ranges of Pliocene era sea level constraints (low vs. high) results in markedly different characterizations of parametric uncertainty and projections of sea level rise over the next five centuries. These results suggest that improved geological data from the Pliocene can help better quantify the model parameters central to marine ice cliff instability (MICI) and improve sea level projections. Second, calibration results are highly sensitive to the choice of prior distributions. Over-constraining the prior distributions (in particular by not allowing very fast cliff disintegration rates), we can mischaracterize parametric uncertainty and drastically underestimate future sea level changes.

2.7.2 Caveats

Our conclusions are subject to the usual caveats that also point to promising and policy relevant research directions. Key methodological caveats include that our calibration approach may not scale well to computer models with long model run times ($>$ 15 minutes), high-dimensional input spaces ($>$ 20 parameters), or a combination of both. For high-dimensional input spaces, our approach would require (1) a large number of particles to sensibly approximate the target distribution; (2) longer mutation stages to move the particles into the high-probability regions; and

(3) prohibitively large amount of computational resources to implement our approach. Our approach may not be suitable for computer models that use multiple processors for a single model run. Selecting an appropriate number of particles remains an open question. Past theoretical work (Crisan and Doucet, 2000) state that using more particles yields better approximations of the target distributions. Here, we set the total particle count with respect to the available resources.

A number of caveats apply to our scientific findings. Using the PSU3D-ICE model at a coarser resolution than previous studies (DeConto and Pollard, 2016; Chang et al., 2016a,b; Pollard et al., 2016) is admittedly a compromise between physical fidelity and run-time feasibility. At coarser resolutions, complex ice processes may not properly coalesce due to the spatial constraints. However, through a simulated example, we found that a single model run at high-resolution (40km) ran nearly eight times longer than one at at coarser resolution (80km). Replicating this calibration study at sharper spatial resolutions (40 km to 10 km) is a natural and worthwhile extension of this study. Promising avenues for future work would include incorporating parallel MCMC approaches such as Multiple-Try Metropolis (Liu et al., 2000) or "emcee" samplers (Goodman and Weare, 2010) to reduce computer model runs in the mutation stage. Finally, the likelihood functions for the paleoclimate records may heavily influence calibration results. We have shown how the choice of expert priors influence calibration, but the influence of likelihood functions remains unexamined.

Acknowledgements

We would like to thank Don Richards, Daniel Gilford, Bob Kopp, Kelsey Ruckert, Vivek Srikrishnans, Robert Ceres, Kristina Rolph, Mahkameh Zarekarizi, and Casey Hegelson for useful discussions. This work was partially supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Earth and Environmental Systems Modeling, MultiSector Dynamics, Contract No. DE-SC0016162 and by the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507. This study was also co-supported by the Penn State Center for CLimate Risk Management. We would like to acknowledge high-

performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Department of Energy, the National Science Foundation, or other funding entities. Any errors and opinions are, of course, those of the authors. We are not aware of any real or perceived conflicts of interest for any authors.

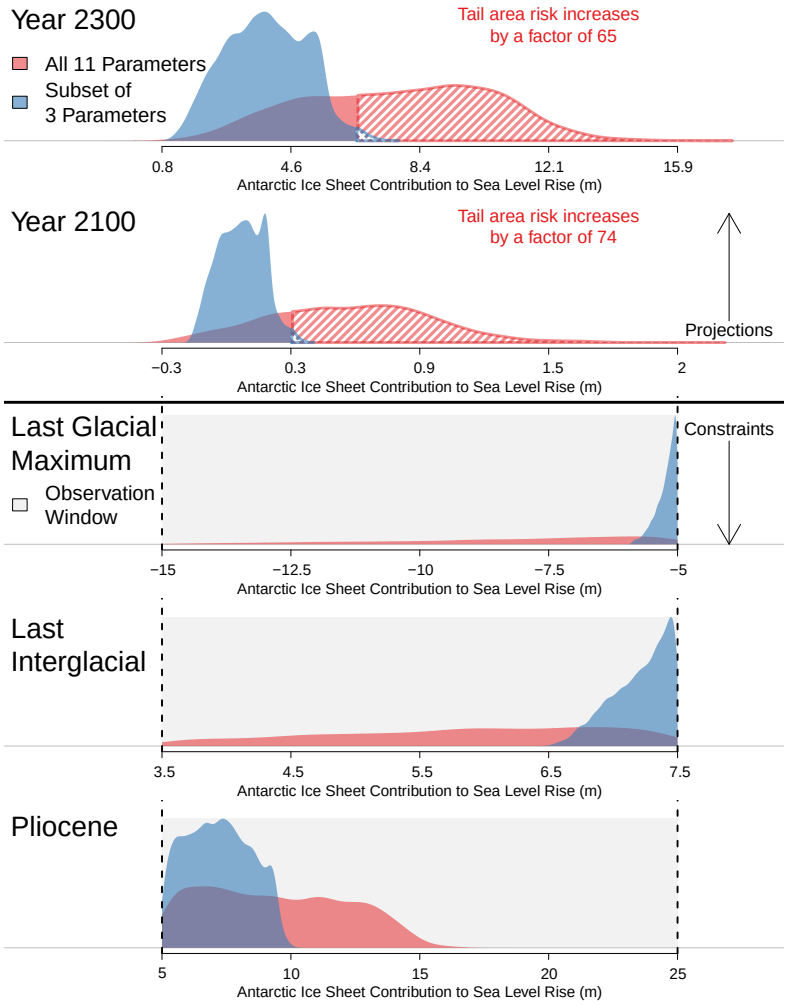

Figure 2.4. Antarctic ice Sheet contribution to sea level rise in the Pliocene (bottom panel), Last Interglacial Age (fourth panel), Last Glacial Maximum (third panel), 2100 (second panel), and 2300 (first panel). Red shading denotes the posterior densities for each time period and projections after calibrating 11 parameters using our fast particle-based approach. Blue shading denotes the posterior densities after calibrating three parameters using emulation-calibration. The light gray shading represents the observational constraints for the Last Glacial Maximum, Last Interglacial Age, and Pliocene. The striped red and striped blue shading represents the 99%th percent quantile for the 11-parameter approach and three-parameter approach, respectively.

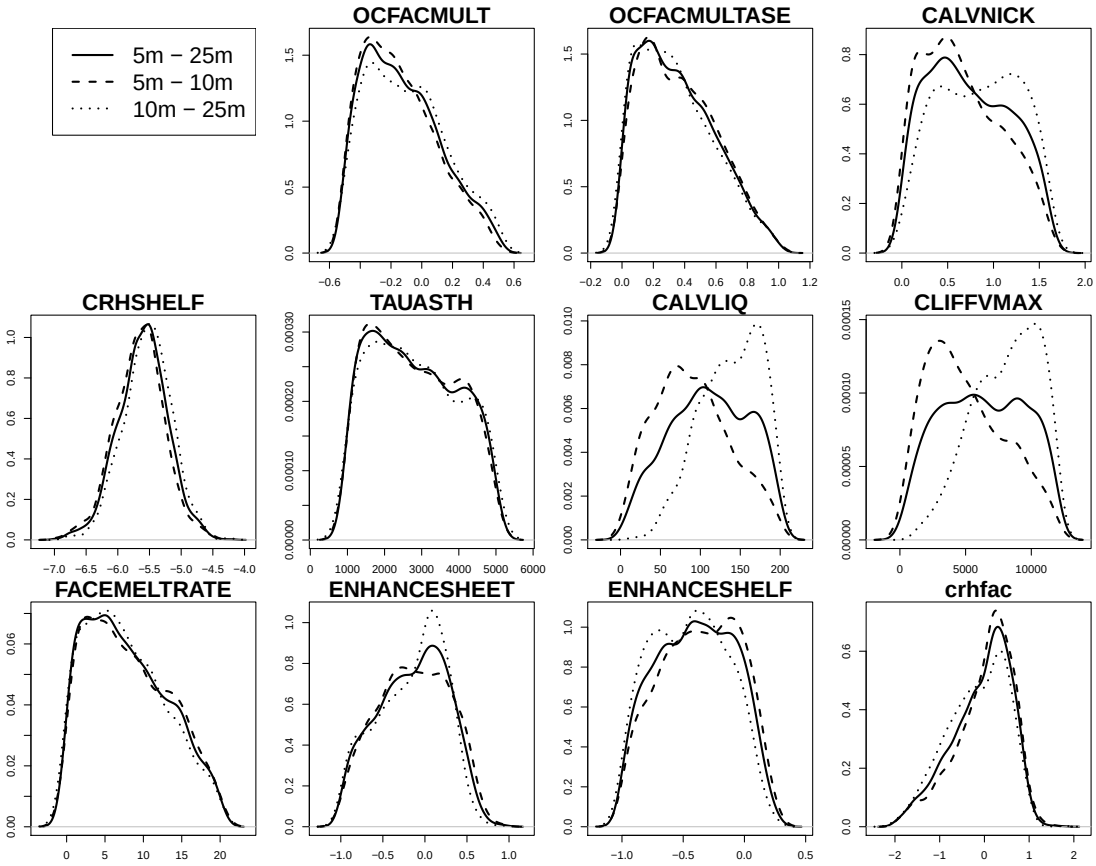

Figure 2.5. Posterior densities of model parameters for calibration using a wide Pliocene window of 5 m to 25 m (solid line), low window of 5 m to 10 m (dashed line), and a high window of 10 m to 25 m (dotted line). There is noticeable change in the densities for three model parameters - CALVNICK, CALVLIQ, and CLIFFVMAX.

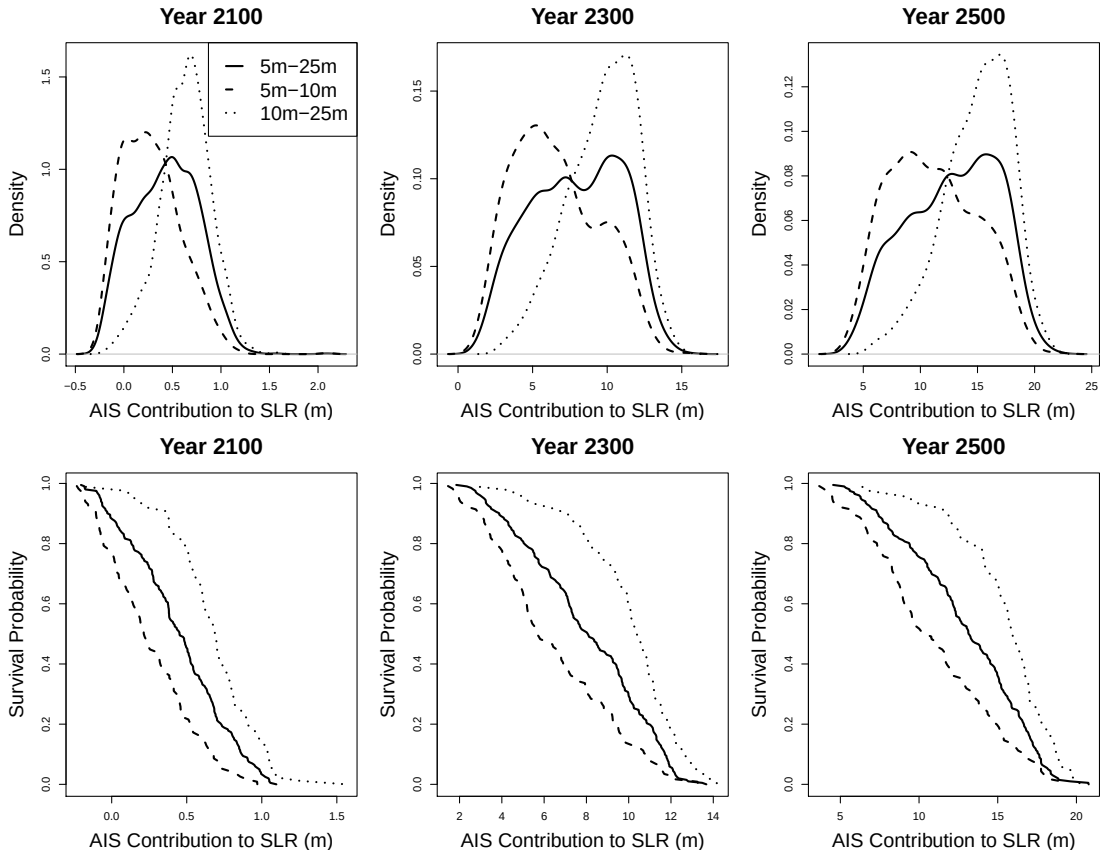

Figure 2.6. (Top Panel) Posterior densities of the projected Antarctic ice sheet's contribution to sea level change in 2100, 2300, and 2500 for calibration using a wide Pliocene window of 5 m to 25 m (solid line), low window of 5 m to 10 m (dashed line), and a high window of 10 m to 25 m (dotted line). (Bottom Panel) Empirical survival function of the projected Antarctic ice sheet's contribution to sea level change in 2100, 2300, and 2500 for calibration using a wide Pliocene window of 5 m to 25 m (solid line), low window of 5 m to 10 m (dashed line), and a high window of 10 m to 25 m (dotted line). constraining the Pliocene windows yield sharper projections of sea level rise. The higher window results in considerably higher projections than the lower window.

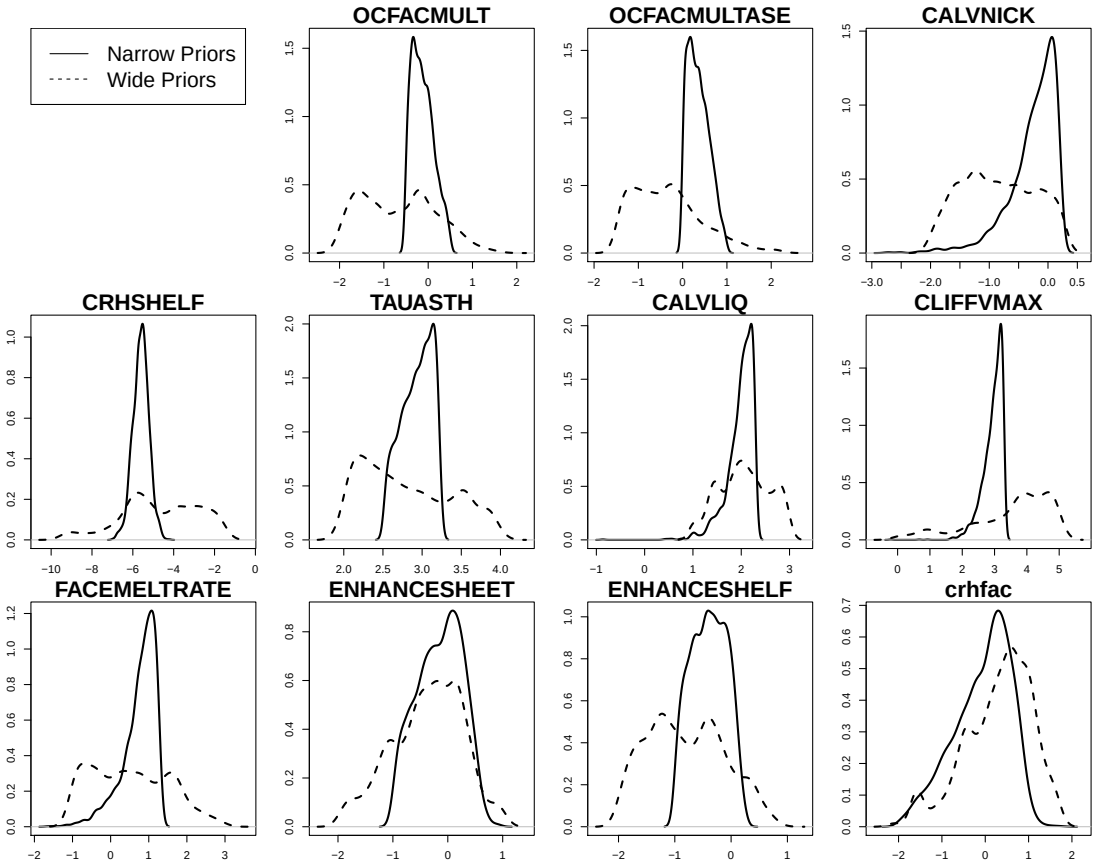

Figure 2.7. Posterior densities of model parameters using expert prior distributions (solid lines) and wider expert prior distributions (dashed lines). The dissimilarity of posterior distributions indicate that calibration results are highly sensitive to the choice of prior distributions.

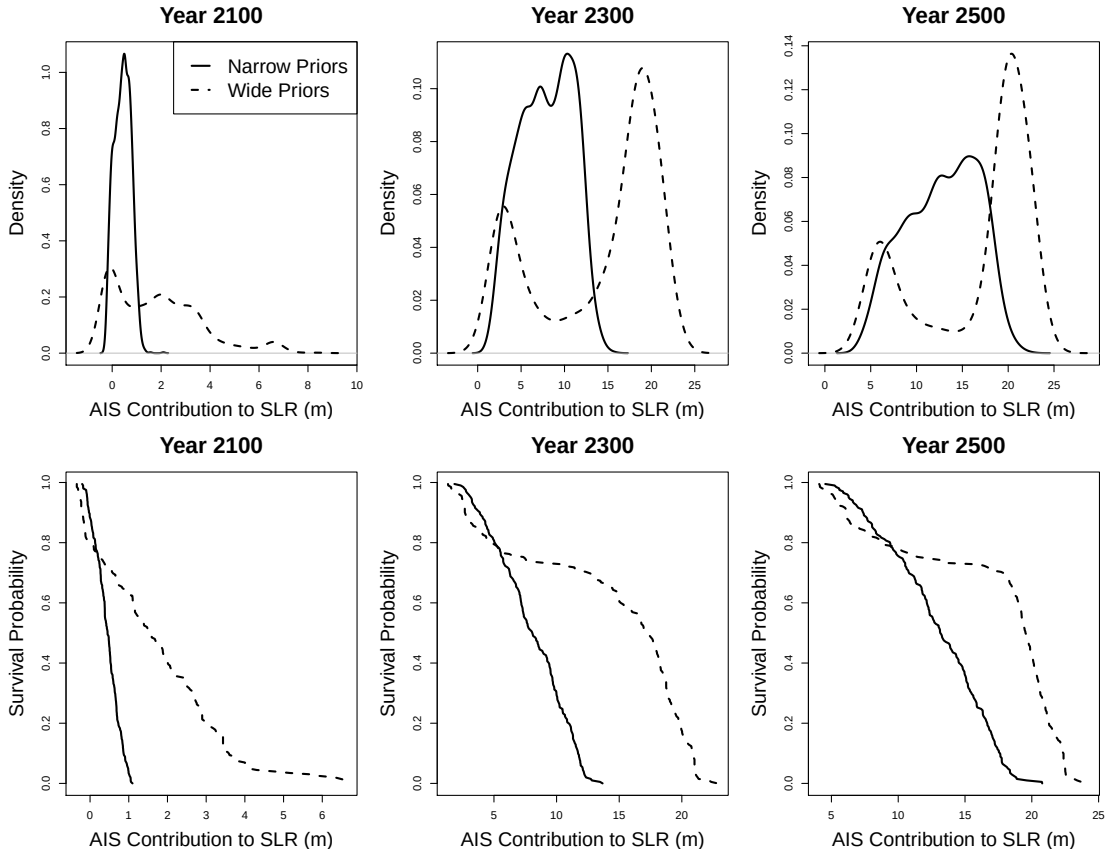

Figure 2.8. (Top Panel) Posterior densities of the projected Antarctic ice sheet's contribution to sea level change in 2100, 2300, and 2500 using expert prior distributions (solid lines) and wider expert prior distributions (dashed lines). (Bottom Panel) Empirical survival function of the projected Antarctic ice sheet's contribution to sea level change in 2100, 2300, and 2500 using expert prior distributions (solid lines) and wider expert prior distributions (dashed lines). For wide prior distributions, projections for future sea level rise is higher and more uncertain, and there exists bi-modality in the projections' posterior predictive distribution.

Chapter 3

PICAR: An Efficient Extendable Approach for Fitting Hierarchical Spatial Models

In this section, I present a projection intrinsic conditional autoregression (PICAR) approach for fitting high-dimensional hierarchical spatial models. This approach is: (1) easily automated; (2) readily extendable to user-specified hierarchical spatial models; and (3) scales well to higher dimensional hierarchical spatial models than is typical of existing methods. The associated manuscript (Lee and Haran, 2019) is currently under review. BSL and MH co-designed the overall study and formulated the statistical method. BSL wrote the computer code and wrote the first draft of the manuscript. MH edited the text.

3.1 Introduction

Hierarchical spatial models are commonly used to model spatial observations across many fields, for example species abundance in ecology, ice presence in glaciology, geo-referenced survey responses in public health studies, and crime incidence in urban areas. A quick search suggests that hierarchical spatial models are featured in thousands of research papers published annually. An important class of hierarchical spatial models is the spatial generalized linear mixed model (SGLMM).

These are flexible models for both point referenced and areal data, where Gaussian random fields are used to model the spatial dependence across locations (Diggle et al., 1998) . Other examples of hierarchical spatial models include spatially varying coefficient processes (Gelfand et al., 2003; Mu et al., 2018), covariate measurement error models (Xia and Carlin, 1998; Bernadinelli et al., 1997; Muff et al., 2015), and co-regionalization models for multivariate responses (Banerjee et al., 2014). Hierarchical spatial models pose considerable computational challenges due to the large number of highly correlated spatial random effects thereby resulting in both slow mixing in Markov Chain Monte Carlo (MCMC) algorithms and costly operations involving large matrices.

In this manuscript, we provide a computationally efficient approach for fitting high-dimensional hierarchical spatial models by decorrelating and reducing the dimensions of the spatial random effects. What sets our projection-based intrinsic conditional autoregression (PICAR) approach apart from existing methods is: (i) our approach to dimension reduction and decorrelation of the random effects is automated; (ii) our approach is easily extendable, that is, it can be easily integrated into a hierarchical modeling scenario using implementations like the probabilistic programming language `stan` (Carpenter et al., 2017); and (iii) our method scales well to higher dimensional hierarchical spatial models than is typical of existing methods. A major advantage of PICAR is that in addition to providing an efficient estimation approach for large datasets, it is easy for non-experts to specify general hierarchical spatial models of their choice in this framework.

Many innovative computational methods have been developed to model high-dimensional spatial data in recent years (cf. Cressie and Johannesson, 2008; Banerjee et al., 2008; Higdon, 1998; Nychka et al., 2015; Lindgren et al., 2011; Katzfuss, 2017; Datta et al., 2016). Comprehensive studies such as Heaton et al. (2019), Bradley et al. (2016), and Sun et al. (2012) examine several of these methods within the context of modeling high-dimensional spatial data. However, these methods primarily focus on linear spatial models with Gaussian observations. A notable exception is the predictive process approach (Banerjee et al., 2008), which easily applies to hierarchical spatial models including SGLMMs. This approach is flexible and efficient, though it requires a careful selection of reference knots which in essence specify a basis (though, see Guhaniyogi et al. (2011) for a point

process-based approach to mitigate some of these challenges). Our approach for selecting a set of basis functions is, in contrast, completely automated and based on the particular spatial data set. For PICAR, the MCMC algorithms mix faster because we use a set of low-dimensional and decorrelated spatial random effects; the predictive process approach random effects tend to be highly dependent. Also, the predictive process requires additional expenses in updating the basis functions at each iteration of the algorithm (roughly $O(np^2 + p^3)$, where p is the number of knots and n is the number of data points); in PICAR, this calculation is avoided and corresponding calculations are linear in p.

Bradley et al. (2019) provides a promising new approach that uses the basis representation while also exploiting conjugate distributions. Computation efficiency comes from the low-dimensional basis representation of the spatial random field. In addition, the conjugate distributions simplify the construction of conditional updates for the MCMC algorithm. However, the full conditional distributions may be difficult to construct as they require computing many matrices, vectors, and constants. There are also open questions regarding the mixing of the resulting Gibbs samplers.

Guan and Haran (2018) use approximate eigendecomposition methods to automatically generate low-rank basis functions; however, the basis functions must be generated iteratively, which adds to the cost of the algorithm. Furthermore, the target distribution changes with each iteration of the algorithm, which prevents the resulting Markov chain from having the theoretical properties assumed in standard MCMC. Data augmentation (Albert and Chib, 1993) has been used to generate Gibbs sampling schemes for the spatial random effects, but this still require costly matrix operations on dense covariance matrices. INLA (Rue et al., 2009; Lindgren et al., 2011) provides a numerical approximation of the posterior distribution. This is an important contribution to computing for hierarchical spatial modeling as it is a very efficient approach and has gained great popularity in recent years. However, while this approach is applicable to a wide array of useful models, it is not easily extendable by non-experts to user-specified hierarchical spatial models. Users can essentially only fit the models that are available to them in their publicly available code. Our approach, in contrast, is easily adaptable to user-specified hierarchical spatial models, as we later demonstrate with our examples in `stan`.

Our method addresses computational challenges by representing the spatial random effects with empirical basis functions. Various basis representations have been directly or indirectly used to model spatial data, for instance in the predictive process approach (Banerjee et al., 2008), random projections (Guan and Haran, 2018, 2019; Banerjee et al., 2013; Park and Haran, 2019), Moran's basis for areal models (Hughes and Haran, 2013), stochastic partial differential equations (Lindgren et al., 2011), kernel convolutions (Higdon, 1998), eigenvector spatial filtering (Griffith, 2003), and multi-resolution basis functions (Nychka et al., 2015; Katzfuss, 2017), among others. We utilize a non-parametric set of basis functions based on the Moran's I statistic and piece-wise linear basis functions. To our knowledge, this is the first approach that readily lends itself to user-specified hierarchical spatial models while also remaining computationally efficient for large datasets. We demonstrate the applicability of PICAR via simulation studies as well as high-dimensional datasets from a forest resource management study and a watershed water quality assessment.

The rest of the paper is organized as follows. In Section 3.2, we describe hierarchical spatial models and discuss their computational challenges. In Section 3.3, we describe our PICAR approach in detail. In Section 3.4, we present simulated examples for: (i) a spatial model for binary observations; (ii) a spatially varying coefficient model for count observations, implemented in PICAR using `stan`; and (iii) a model for ordered categorical spatial data that cannot be fit using existing publicly available code but can be easily fit using PICAR. In Section 3.5 we apply PICAR to two large spatial datasets: occurrence of a parasitic species of dwarf mistletoe in Minnesota and water quality ratings in Maryland watersheds. Finally, we provide a summary and directions for future research in Section 3.6.

3.2 Hierarchical Spatial Models

We begin by describing a general framework for hierarchical spatial models and provide several examples that will be explored in depth via simulation studies. We also provide a general discussion of the computational challenges for fitting these models.

3.2.1 Model Specification

Let $Z(s)$ denote a spatial process at location s in a spatial domain $\mathcal{D} \subset \mathbb{R}^d$ where d is typically 2 or 3. We define $Z(s)$ as:

$$Z(s) = X(s)\beta + w(s) + \epsilon(s), \text{ for s} \in D, \tag{3.1}$$

where $X(s)$ is a set of k covariates associated with location s and β is a k-dimensional vector of coefficients. The micro-scale measurement errors or nugget are modeled as an uncorrelated Gaussian process with zero mean and variance τ^2 where $\epsilon(s) \sim N(0, \tau^2)$ for all $s \in \mathcal{D}$.

We impose spatial dependence by modeling the spatial random effects $\mathbf{W} = \{w(s) : s \in \mathcal{D}\}$ as a stationary zero-mean Gaussian process with a positive definite covariance function $C(\cdot)$. For a finite set of locations $s = (s_1, ..., s_n)$, the spatial random effects $\mathbf{W}$ are distributed as a multivariate normal distribution $\mathbf{W}|\Theta \sim N(0, C(\Theta))$ with covariance function parameters Θ and the covariance matrix $C(\Theta)$ where $C(\Theta)_{ij} = \text{cov}(w(s_i), w(s_j))$. The Matérn covariance function is a widely used class of stationary and isotropic covariance functions (Stein, 2012) with parameters $\Theta = (\sigma^2, \phi, \nu)$ such that:

$$C(s_i, s_j) = \sigma^2 \frac{1}{\Gamma(\nu)2^{\nu-1}} \Big(\sqrt{(}2\nu)\frac{h}{\phi}\Big)^{\nu} K_\nu\Big(\sqrt{(}2\nu)\frac{h}{\phi}\Big),$$

where $R(\phi)$ is the correlation matrix, $h = ||s_i - s_j||$ is the Euclidean distance between locations s_i and s_j, $\sigma^2 > 0$ is the partial sill or scale parameter of the process, and $\phi > 0$ is the range parameter for spatial dependence. $K_\nu(\cdot)$ is the modified Bessel function of the second kind where the smoothness parameter ν is commonly fixed prior to model fitting.

Hierarchical spatial models may be broadly described as (cf. Wikle et al., 1998):

Data Model: $Z(s)|\beta, \mathbf{W}, \tau^2 \sim N(X(s)\beta + w(s), \tau^2)$

Process Model: $\mathbf{W}|\phi, \sigma^2 \sim N(0, \sigma^2 R_\phi)$

Parameter Model: $\beta \sim p(\beta), \phi \sim p(\phi), \sigma^2 \sim p(\sigma^2), \tau^2 \sim p(\tau^2)$

3.2.2 Examples of Hierarchical Spatial Models

Here we provide examples of hierarchical spatial models. The first is the class of spatial generalized linear mixed models for non-Gaussian data, the second is a spatially varying coefficient model, and the third is a cumulative-logit model for ordered categorical data.

Spatial Generalized Linear Mixed Models

Non-Gaussian spatial observations are typically modeled using spatial generalized linear mixed models (SGLMMs) (Diggle et al., 1998). Let $\{Z(s) : s \in \mathcal{D}\}$ be a non-Gaussian spatial random field. Assuming $Z(s)$ are conditionally independent given the latent random spatial field $\mathbf{W}$, the conditional mean $E[Z(s)|\beta, \mathbf{W}, \epsilon(s)]$ can be modeled through a linear predictor $\eta(s)$:

$$\eta(s) = g\{E[Z(s)|\beta, \mathbf{W}], \epsilon(s)\} = X(s)\beta + w(s) + \epsilon(s),$$

where $g(\cdot)$ is a known link function. Binary and count observations are two common types of non-Gaussian spatial data, and these can be modeled using the binary SGLMM with logit link and the Poisson SGLMM with log link, respectively.

The general Bayesian hierarchical framework for non-Gaussian spatial observations is:

Data Model:	$Z(s)\|\eta(s) \sim f(\eta(s))$
	$\eta(s) = g(\mathbb{E}[Z(s)\|\beta, \mathbf{W}), \epsilon(s)]) = X(s)\beta + w(s) + \epsilon(s)$
Process Model:	$\mathbf{W}\|\phi, \sigma^2 \sim N(0, \sigma^2 R_\phi)$
	$\epsilon(s)\|\tau^2 \sim N(\mathbf{0}, \tau^2)$
Parameter Model:	$\beta \sim p(\beta), \phi \sim p(\phi), \sigma^2 \sim p(\sigma^2), \tau^2 \sim p(\tau^2)$

Spatially Varying Coefficient Models

Spatially varying coefficient models (Gelfand et al., 2003) consider cases where the fixed effects β in Equation 3.1 vary across space. For the case with a single predictor $X(s)$, the data model is $Z(s) = \beta_0 + \beta_1 X(s) + \beta_1(s)X(s) + w(s) + \epsilon(s)$, where β_0 is the intercept, β_1 is the fixed effect, $\beta_1(s)$ is the spatially varying coefficient term, and $w(s)$ and $\epsilon(s)$ are the same as in Equation 3.1. Here, $\mathbf{B} = (\beta_1(s_1), ..., \beta_1(s_n))$

is the n-dimensional vector of spatially varying coefficients, and $\mathbf{B} \sim N(0, \sigma_\beta^2 R_{\phi_\beta})$ where σ_β^2 is the partial sill and ϕ_β is the range parameter for the spatial random process $\mathbf{B}$.

For cases with k predictors, we have the following hierarchical spatial model:

$$
\begin{aligned}
&\textbf{Data Model:} && Z(s)|\eta(s) \sim f(\eta(s)) \\
& && \eta(s) = X(s)\beta + X(s)\beta(s) + w(s) + \epsilon(s) \\
&\textbf{Process Model:} && (\mathbf{W}, \mathbf{B})^T|\phi, \mathbf{T} \sim \mathcal{N}(\mathbf{0}, R_\phi \otimes \mathbf{T}) \\
& && \epsilon(s)|\tau^2 \sim N(\mathbf{0}, \tau^2) \\
&\textbf{Parameter Model:} && \beta \sim \pi(\beta), \quad \tau^2 \sim \pi(\tau^2), \quad \phi \sim \pi(\phi), \quad \mathbf{T} \sim \pi(\mathbf{T})
\end{aligned}
$$

where β is the k-dimensional vector of the fixed effects, $\beta(s) = (\beta_1(s), ..., \beta_k(s))$ is a k-dimensional vector of the spatially varying coefficients for location s, $\mathbf{B} = (\beta(s_1), ..., \beta(s_n))$ is the nk-dimensional vector of all spatially varying coefficients, $\mathbf{W} = (W(s_1), ..., W(s_n))$ is the n-dimensional vector of the spatial random effects, R_ϕ and τ^2 are the correlation matrix and nugget variance described in Section 3.2.1, and $\mathbf{T}$ is a $(k+1) \times (k+1)$ positive definite matrix.

Cumulative-Logit Models for Ordinal Spatial Data

Ordered categorical (ordinal) data are categorical responses with a natural ordering, and commonly used in survey questionnaires, patient responses in clinical trials, and quality assurance ratings for industrial processes. (Higgs and Hoeting, 2010; Schliep and Hoeting, 2013) develop a hierarchical spatial model for ordinal data. In this study, we examine the proportional-odds cumulative logit model (Agresti, 2010) for ordered categorical data. Let $Z(s)$ be the observations at location $s \in \mathcal{D}$ with J ordered categories. Note that each ordered category corresponds to a probability $\pi(s) = \{\pi_1(s), \pi_2(s), ..., \pi_J(s)\}$, where $\pi_i(s) = \Pr(Z(s) = i)$ for $i = 1, ..., J$. Here, we consider $J - 1$ cumulative probabilities denoted as $\gamma_j(s) = P(Z(s) \leq j) = \pi_1(s) + ... + \pi_j(s)$. The cumulative logit is defined as:

$$
\log\left(\frac{P(Z(s) \leq j)}{1 - P(Z(s) \leq j)}\right) = \log\left(\frac{\gamma_j(s)}{1 - \gamma_j(s)}\right) = \theta_j - X(s)\beta - w(s) - \epsilon(s),
$$

where θ_j is the intercept or "cutoff" for the j-th category, $X(s)$, β, $w(s)$ and $\epsilon(s)$ are the same as in Equation 3.1. The model for the cumulative probabilities γ_j is:

$$\gamma_j(s) = P(Z(s) \leq j) = \frac{\exp\{\theta_j - (X(s)\beta + w(s) + \epsilon(s))\}}{1 + \exp\{\theta_j - (X(s)\beta + w(s) + \epsilon(s))\}}.$$

Consequently, the probabilities for the individual J categories are:

$$P(Z(s) = j) = \begin{cases} \gamma_1(s), & j = 1 \\ \gamma_j(s) - \gamma_{j-1}(s), & 2 \leq j \leq J-1 \\ 1 - \gamma_{J-1}(s), & j = J \end{cases}$$

To avoid identifiability issues, we typically fix the first cutoff to be $\theta_1 = 0$ (Johnson and Albert, 2006). Note that the θ_j's are constrained by the ordering $\theta_j > \theta_k$ for $j > k$. Through a transformation (Higgs and Hoeting, 2010; Albert and Chib, 1997), we can generate unconstrained cutoff parameters $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{J-1})$, where $\alpha_1 = -\infty$, $\alpha_2 = \log(\theta_2)$, and $\alpha_j = \log(\theta_j - \theta_{j-1})$ for $j = 3, ..., J-1$. The inverse transformation is $\theta_j = \sum_{i=1}^{J-1} \exp\{\alpha_i\}$. The hierarchical spatial model framework is as follows:

Data Model:
$$Z(s)|\gamma(s) \sim f(\gamma(s))$$
$$\gamma_j(s)|\beta, \theta, \mathbf{W}, \epsilon(s) = \frac{\exp\{\theta_j - (X(s)\beta + w(s) + \epsilon(s))\}}{1 + \exp\{\theta_j - (X(s)\beta + w(s) + \epsilon(s))\}}$$
$$\theta_j|\alpha = \sum_{i=1}^{J-1} \exp\{\alpha_i\}$$

Process Model:
$$\mathbf{W}|\phi, \sigma^2 \sim N(\mathbf{0}, \sigma^2 R_\phi)$$
$$\epsilon(s)|\tau^2 \sim N(\mathbf{0}, \tau^2)$$

Parameter Model:
$$\alpha \sim p(\alpha), \quad \beta \sim p(\beta), \quad \phi \sim p(\phi),$$
$$\sigma^2 \sim p(\sigma^2), \quad \tau^2 \sim p(\tau^2)$$

3.2.3 Model Fitting and Computational Challenges

In hierarchical spatial models, computational challenges are rooted in the dimensionality and correlation of the spatial random effects $\mathbf{W}$. Hierarchical spatial models typically require a costly evaluation of an $n-$dimensional multivariate normal likelihood function ($\mathcal{O}(n^3)$) at each iteration of the MCMC algorithm. Moreover, highly correlated spatial random effects can lead to poor mixing in MCMC

algorithms (cf. Christensen et al., 2006; Haran et al., 2003).

There is a large literature on addressing computational challenges in spatial models though the vast majority of methods are focused on linear Gaussian spatial models where the latent spatial variables do not need to be integrated out. Popular approaches include low-rank approximations (Cressie and Johannesson, 2008; Banerjee et al., 2008), compact support or covariance tapering (Furrer et al., 2006; Stein, 2013), multiresolution approaches (Nychka et al., 2015; Katzfuss, 2017), and sparse representations of the $n \times n$ precision matrix via spatial partial differential equations (Lindgren et al., 2011) or nearest-neighbor Gaussian processes (Datta et al., 2016). These typically focus on the marginal distribution of the spatial observations $\mathbf{Z} = (Z(s_1, ..., Z(s_n))$, where $\mathbf{Z}|\beta, \sigma^2, \phi, \tau^2 \sim \mathcal{N}(X\beta, \Sigma(\sigma^2, \phi, \tau^2))$. Since these methods are built on integrating out the spatial random effects in closed form, they are not easily extended to more complex hierarchical spatial models.

For hierarchical spatial models with non-Gaussian observations, Sengupta and Cressie (2013) and Sengupta et al. (2016) extend fixed-rank kriging (Cressie and Johannesson, 2008) to non-Gaussian satellite imagery by: (1) representing the spatial random effects using bi-square basis functions; (2) estimating the model parameters via the expectation-maximation (EM) algorithm; and (3) embedding Laplace approximations in the E-step to improve computational efficiency. The predictive process approach (Banerjee et al., 2008) also implement a basis representation of the spatial random effects $W(s) \in \mathbb{R}^n$. Prior to model fitting, m reference locations or knots $s^* = \{s_1^*, s_1^*, ..., s_m^*\}$ are selected where $m \ll n$. Predictive processes approximates the spatial random effects such that $W(s) \approx C(s, s^*)C^{*-1}W(s^*)$, where $W(s^*)$ are the realizations of the Gaussian random field at knot locations s^*, $C^* = C(s^*, s^*)$ represents the $m \times m$ covariance matrix corresponding to the knots, and $C(s, s^*)$ denotes the cross-covariance between the observed locations (s) and the knot locations (s^*). Knots can be selected using an adaptive approach based on point processes (Guhaniyogi et al., 2011). Note that the computational speedup comes from utilizing a lower-dimensional set of spatial random effects $W(s^*)$ and an $m \times m$ covariance matrix. However, predictive process must construct the basis functions matrix $C(s, s^*)C^{*-1}$ at each iteration of the algorithm, which incurs a cost of roughly $O(m^3 + nm^2)$.

Guan and Haran (2018) use random projections to generate approximate eigen-

vector basis functions. These basis functions are linked to the Matérn class class of covariance functions, and the dominant computational cost is driven by large matrix-to-matrix multiplications, which can be easily parallelized. The spatial random effects are reparameterized and approximated as $W(s) \approx \tilde{U}_m \tilde{D}_m^{1/2} \delta$, where $\tilde{U}_m$ and $\tilde{D}_m$ are the first m approximate eigencomponents of the covariance matrix $\sigma^2 R_\phi$ for SGLMMs and δ are the m-dimensional reparamterized spatial random effects. While this approach bypasses knot selection, it still requires repeated constructions of the approximate eigenvector basis functions with a cost of $O(m^3 + nm^2)$.

Re-parameterization approaches (Christensen et al., 2006; Haran et al., 2003; Guan and Haran, 2018) decorrelate the spatial random effects, which often results in faster mixing MCMC algorithms. However, these techniques can be very expensive when data are on the order of thousands of data points since the reparameterization step itself can be expensive. Data augmentation approaches (De Oliveira, 2000; Albert and Chib, 1993) apply to some classes of hierarchical models, resulting in a Gibbs sampler for the spatial random effects, but this still requires large matrix operations on dense covariance matrices, and does not necessarily address mixing issues in the resulting MCMC algorithm.

3.3 PICAR Approach

In this section, we present our projection-based intrinsic conditional autoregression (PICAR) approach that is designed to efficiently fit hierarchical spatial models. In this framework, we represent spatial random effects $\mathbf{W} = (W(s_1), ..., W(s_n))$ as a linear combination of basis functions:

$$\mathbf{W} \approx \mathbf{\Phi}\delta \qquad , \qquad \delta \sim \mathcal{N}(0, \Sigma_\delta),$$

where $\mathbf{\Phi}$ is an $n \times p$ basis function matrix where each column denotes a basis function, $\delta \in \mathbb{R}^p$ are the re-parameterized spatial random effects (or basis coefficients), and Σ_δ is the $p \times p$ covariance matrix for the weights. Basis functions can be interpreted as a set of distinct spatial patterns that can be used to construct a spatial random field, along with their coefficients. Basis representation

has been a popular approach to model spatial data (cf. Cressie and Johannesson, 2008; Banerjee et al., 2008; Hughes and Haran, 2013; Lindgren et al., 2011; Rue et al., 2009; Christensen et al., 2006; Haran et al., 2003; Griffith, 2003; Higdon, 1998; Nychka et al., 2015). Examples of basis functions include splines, wavelets, empirical orthogonal functions, combinations of sines and cosines, piece-wise linear functions, and many others. Basis representations tend to be computationally efficient as they help bypass large matrix operations, reduce the dimensions of the spatial random effects, and as in our case, decorrelate the spatial random effects $\mathbf{W}$.

3.3.1 PICAR Approach

The Projection-based intrinsic conditional auto-regression (PICAR) approach can be outlined as follows:

1. Generate a triangular mesh on the spatial domain $\mathcal{D} \subset \mathbb{R}^2$.
2. Construct a spatial field on the mesh vertices using non-parametric basis functions.
3. Interpolate onto the observation locations using piece-wise linear basis functions.

We provide details for each step below.

Mesh Construction

Prior to fitting the model, we generate a mesh enveloping the observed spatial locations via Delaunay Triangulation (Hjelle and Dæhlen, 2006). Here, we divide the spatial domain D into a collection of non-intersecting irregular triangles. The triangles can share a common edge, corner (i.e. nodes or vertices), or both. The mesh generates a latent undirected graph $G = \{V, E\}$, where $V = \{1, 2, ..., m\}$ are the mesh vertices and E are the edges. Each edge E is represented as a pair (i, j) denoting the connection between i and j. The graph G is characterized by its weights matrix $\mathbf{N}$, an $m \times m$ matrix where $N_{ii} = 0$ and $N_{ij} = 1$ when mesh node i is connected to node j and $N_{ij} = 0$ otherwise. The triangular mesh is built using

the **R-INLA** package (Lindgren et al., 2015). Guidelines for mesh construction are provided in Lindgren et al. (2015), and details pertaining to algorithms for Delaunay triangulation can be found in Hjelle and Dæhlen (2006).

Moran's Basis Functions

We generate a spatial random field on the set of mesh vertices V of graph G using the Moran's basis functions (Hughes and Haran, 2013; Griffith, 2003). Griffith (2003) proposes an augmented spatial generalized linear mixed model using a subset of eigenvectors of the Moran's operator $(\mathbf{I} - \mathbf{1}\mathbf{1}'/m)\mathbf{W}(\mathbf{I} - \mathbf{1}\mathbf{1}'/m)$, where $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is a vector of 1's. Note that this operator is a component of the Moran's I statistic:

$$I(A) = \frac{m}{\mathbf{1}'\mathbf{W}\mathbf{1}} \frac{\mathbf{Z}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/m)\mathbf{W}(\mathbf{I} - \mathbf{1}\mathbf{1}'/m)\mathbf{Z}}{\mathbf{Z}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/m)\mathbf{Z}},$$

a diagnostic of spatial dependence (Moran, 1950) used for areal spatial data. Values of the Moran's I above $-\frac{1}{m-1}$ indicate positive spatial autocorrelation and values below $-\frac{1}{m-1}$ indicate negative spatial autocorrelation (Griffith, 2003). Positive eigencomponents of the Moran's operator correspond to varying magnitudes and patterns of positive spatial dependence, or clustering. For the triangular mesh, the positive eigenvectors represent the patterns of spatial dependence among the mesh nodes, and their corresponding eigenvalues denote the magnitude of spatial dependence. Figure 3.1 illustrates the first 25 eigenvectors of the Moran's operator.

We construct the Moran's basis function matrix $\mathbf{M} \in \mathbb{R}^{m \times p}$, by selecting the first p eigenvectors of the Moran's operator where $p \ll m$. In Section 3.3.3, we provide an automated heuristic for selecting a suitable rank p. We can generate a spatial random field on the mesh vertices by taking linear combinations of the Moran's basis functions (contained in matrix $\mathbf{M}$) and their corresponding weights $\delta \in \mathbb{R}^p$. In Section 3.3.2, we provide a general framework for estimating δ in hierarchical spatial models.

Piece-wise Linear Basis Functions

To complete the PICAR approach, we introduce a set of piece-wise linear basis functions (Brenner and Scott, 2007) to interpolate points within the triangular

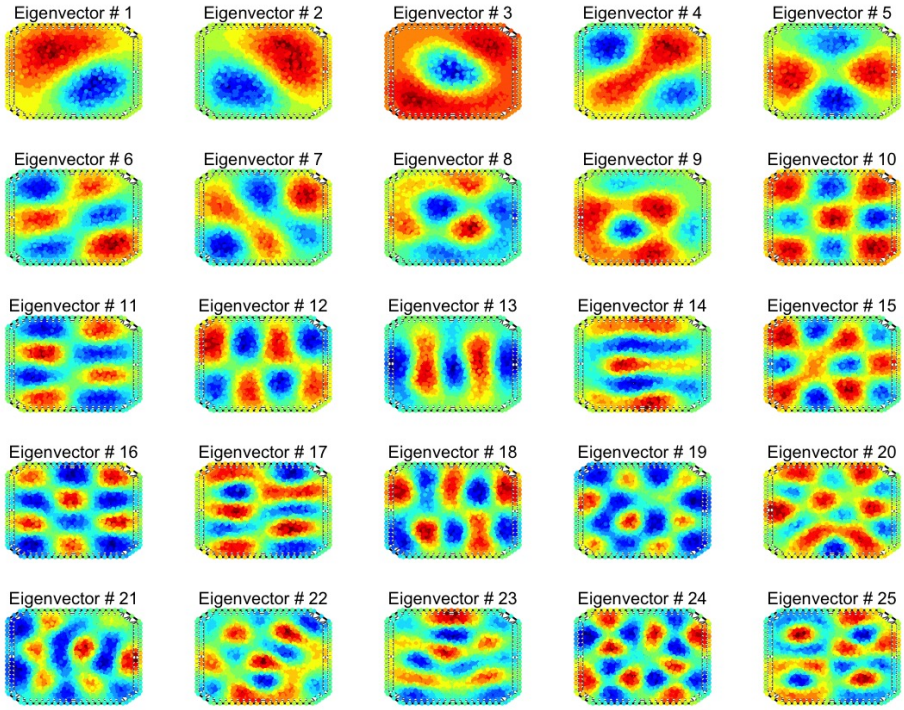

Figure 3.1. The leading 25 eigenvectors of the Moran's operator generated on the triangular mesh. The distinct spatial patterns construct the latent spatial random field for hierarchical spatial models.

mesh (i.e. the undirected graph $G = (V, E)$). We construct a spatial random field on the mesh nodes $\tilde{\mathbf{W}} = (W(v_1), ..., W(v_m))$ where $v_i \in V$ and then project, or interpolate, onto the observed locations $\mathbf{W} = (W(s_1), ..., W(s_n))$ where $s_i \in \mathcal{D}$. The latent spatial random field $\mathbf{W}$ can be represented as $\mathbf{W} = \mathbf{A}\tilde{\mathbf{W}}$, where $\mathbf{A}$ is an $n \times m$ projector matrix containing the piece-wise linear basis functions.

The rows of $\mathbf{A}$ correspond to an observation location $s_i \in D$, and the columns correspond to a mesh node $v_i \in V$. The ith row of $\mathbf{A}$ contains the weights to linearly interpolate $W(s_i)$. To illustrate, when the observation location s_i is wholly contained within one of the mesh triangles, there will be three non-zero values in the ith row of the projector matrix $\mathbf{A}$, each corresponding to a mesh node $v_j \in V$. When the observation location lies on an edge between two mesh nodes, there will be two non-zero values in the corresponding row of $\mathbf{A}$. Finally, there will only be one non-zero value in the corresponding row when the observation location and mesh node share the same location. In practice, we use an $n \times m$ projector matrix

A for fitting the hierarchical spatial model. For model validation and prediction, we generate an $n_{CV} \times m$ projector matrix $\mathbf{A}_{CV}$ that interpolates onto the n_{CV} validation locations.

In Figure 3.2, we demonstrate how the piece-wise linear basis functions can interpolate an observation location that is wholly contained in a triangle. Here, point D is the observation location, points A, B, and C are the triangle vertices, and π_1, π_2, and π_3 are the weights, where $\sum_i^3 \pi_i = 1$. π_1 is the proportion of the area of the triangle opposite of vertex A to the entire triangle. The same holds for values π_2 and π_3 with corresponding vertices B and C, respectively. We interpolate point D as the weighted mean of the three triangle vertices where $D \approx \pi_1 A + \pi_2 B + \pi_3 C$.

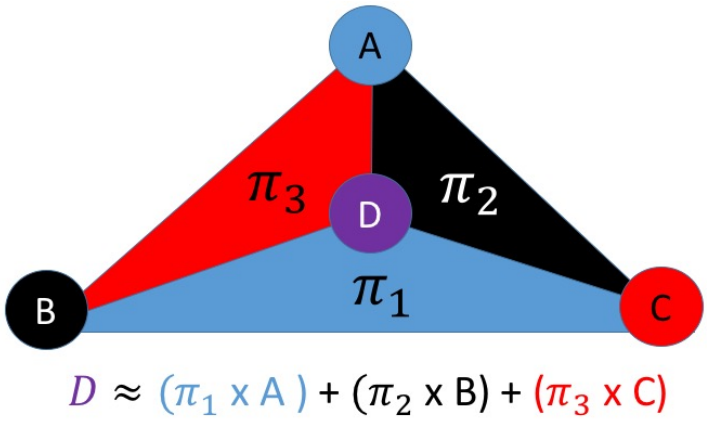

Figure 3.2. Diagram of the piece-wise linear basis functions. Point D is the observation location, points A, B, and C are the triangle vertices, and π_1, π_2, and π_3 are the corresponding weights. The weights π_1, π_2, and π_3 correspond to the proportion of the area of the specified triangle to the area of the larger triangle. We interpolate point D by taking the weighted mean of the three triangle vertices where $D \approx \pi_1 A + \pi_2 B + \pi_3 C$

3.3.2 Bayesian Hierarchical Spatial Model using PICAR

In the previous section, we introduced three major components of the PICAR approach: (1) the Moran's basis function matrix $\mathbf{M} \in \mathbb{R}^{m \times p}$; (2) the projector matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$; and (3) the corresponding weights $\delta \in \mathbb{R}^p$. Given a set of weights δ and the Moran's basis functions $\mathbf{M}$, we can build a spatial random field on the triangular mesh nodes $v \in V$ as $\tilde{\mathbf{W}} = \mathbf{M}\delta$, where $\tilde{\mathbf{W}} = (W(v_1), ..., W(v_m))$ for $v_i \in V$. Next, we linearly interpolate the latent spatial random field at the

observation locations as $\mathbf{W} = \mathbf{A}\tilde{\mathbf{W}} = \mathbf{AM}\delta$, where $\mathbf{W} = (W(s_1), ..., W(s_n))$ for $s_i \in \mathcal{D}$, the spatial domain. An overview of these operations is provided in Figure 3.3.

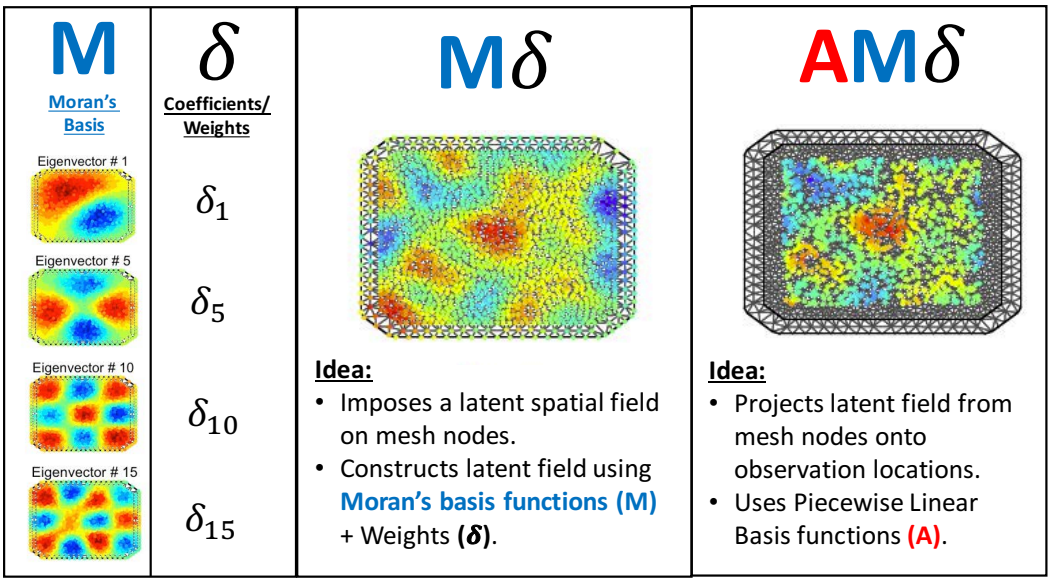

Figure 3.3. Diagram of the basis functions within the PICAR framework. The Moran's basis functions (left) represent distinct spatial patterns, and the coefficients (δ) denote the associated weights. The operation $\mathbf{M}\delta$ constructs a latent field on the mesh nodes. The operation $\mathbf{AM}\delta$ projects the mesh nodes onto the observation locations and generates a spatial random field.

The PICAR approach can be embedded into the hierarchical spatial model framework:

$$
\begin{aligned}
&\textbf{Data Model:} && Z(s)|\eta(s) \sim f(\eta(s)), \\
& && \eta(s) = g(E[Z(s)|\beta, \delta]) = X(s)\beta + [\mathbf{AM}\delta](s), \\
&\textbf{Process Model:} && \delta|\tau \sim \mathcal{N}(0, \tau^{-1}(\mathbf{M}'\mathbf{Q}\mathbf{M})^{-1}), \\
&\textbf{Parameter Model:} && \beta \sim N(\mu_\beta, \Sigma_\beta), \quad \tau \sim G(\alpha_\tau, \beta_\tau),
\end{aligned}
$$

where $\mathbf{A}$ is the projector matrix, $\mathbf{M}$ is the Moran's basis functions matrix, δ are the basis coefficients, $\mathbf{Q}$ is the prior precision matrix for the mesh vertices, τ is the precision parameter, and $\alpha_\tau, \beta_\tau, \mu_\beta$, and Σ_β are the hyperparameters.

By default, we set $\mathbf{Q}$ to be the precision matrix of an intrinsic conditional auto-regressive model (ICAR) fit on the mesh vertices V. Here, $\mathbf{Q} = (\text{diag}(\mathbf{N1}) -$

$\mathbf{N}$), where $\mathbf{N}$ is the adjacency or weight matrix from Section 3.3.1 and $\mathbf{1}$ is m-dimensional vector of 1s. Since $\mathbf{Q}$ is not positive definite, this framework cannot be used within the likelihood function; however, it can be set as the prior distribution for the spatial random effects as part of the Bayesian hierarchical spatial model (Besag et al., 1991). We introduce alternative precision matrices in Section 3.3.3 and provide a comparative analysis across matrices in Section 3.4.

3.3.3 Automating PICAR

The traditional hierarchical spatial model (Section 3.2.1) assumes that the true latent spatial random field $\mathbf{W} = \{W(s_1), W(s_2), ..., W(s_n)\}$ is a Gaussian process such that $\mathbf{W} \sim N(0, \sigma^2 R_\phi)$ with partial sill σ^2 and correlation matrix R_ϕ. On the other hand, the PICAR approach considers the latent spatial random field following a basis representation such that $\mathbf{W} \approx \mathbf{AM}\delta$, where $\delta \sim N(0, \tau^{-1}(\mathbf{M'QM})^{-1})$, $\mathbf{M}$ is the $m \times p$ Moran's basis function matrix, and $\mathbf{A}$ is the $n \times m$ projector matrix. An alternative formulation of the latent spatial random field is

$$\mathbf{W} \sim N(0, \tau^{-1}\mathbf{AM}(\mathbf{M'QM})^{-1}\mathbf{M'A'})$$

Our objective is to accurately represent the true latent state using PICAR's basis representation. To that end, we can tune the rank of the Moran's operator rank($\mathbf{M}$) and the prior precision matrix $\mathbf{Q}$ of the mesh vertices.

The following automated heuristic selects an appropriate rank for the Moran's basis. First, we generate a set $\mathcal{P}$ consisting of h equally spaced points within the interval $[2, P]$ where P is the maximum rank and h is the interval resolution ($h = P - 1$ by default). Here, $P < m$ and both P and h are chosen by the user. For each $p \in \mathcal{P}$, we construct an $n \times (k + p)$ matrix of augmented covariates $\tilde{X} = [X \quad \mathbf{AM_p}]$ where $X \in \mathbb{R}^{n \times k}$ is the original covariate matrix, $\mathbf{A} \in \mathbb{R}^{n \times m}$ is the projector matrix, and $\mathbf{M_p} \in \mathbb{R}^{m \times p}$ are the leading p eigenvectors of the Moran's operator. Next, we use maximum likelihood approaches to fit the appropriate generalized linear model (GLM) for the response type (e.g. binary, count, or ordered categorical). Finally, we select the rank p that yields the lowest out-of-sample cross-validated mean squared prediction error (CVMSPE). Figure 3.4 illustrates the automated heuristic for a binary dataset (n=1000). Note that rank

$p = 62$ results in the lowest CVMSPE.

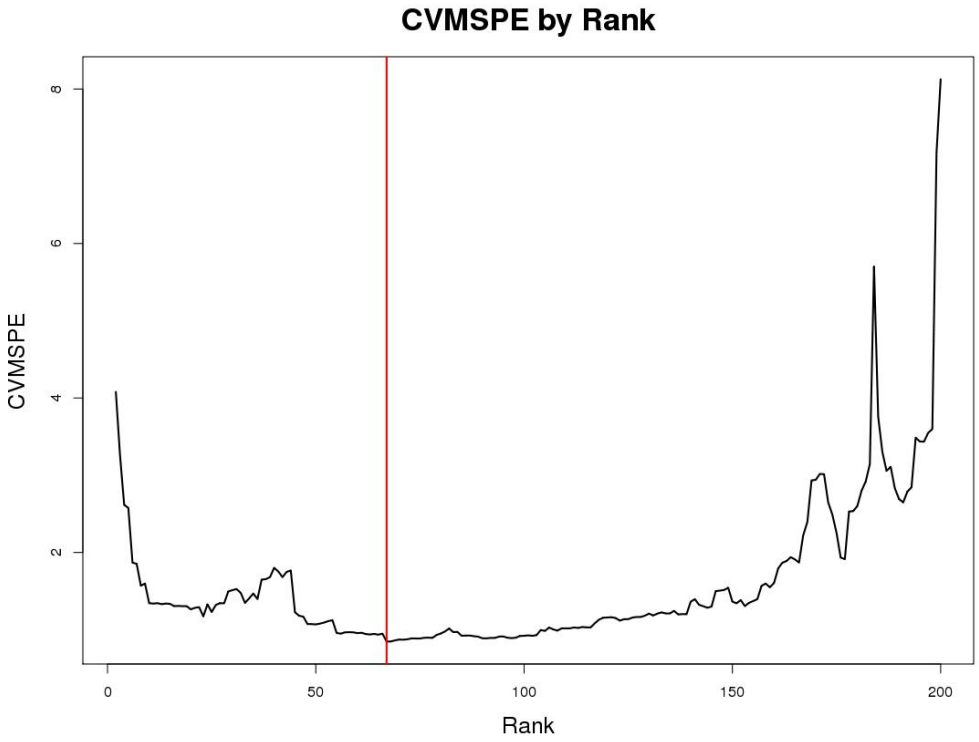

Figure 3.4. Cross-validated mean squared prediction error (CVMSPE) for ranks 1-200 using the automated heuristic. The vertical red line denotes the chosen rank ($p = 68$) with lowest CVMPSE.

Next, we provide some choices for $\mathbf{Q}$, the prior precision matrix for the mesh vertices $\tilde{\mathbf{W}}$. By default (Section 3.3.1), we set $\mathbf{Q}$ to be the precision matrix of an intrinsic conditional auto-regressive model (ICAR). Similarly, we could set $\mathbf{Q}$ as the precision matrix of a conditional auto-regressive model (CAR). Here, $\mathbf{Q} = (\mathbf{N1} - \rho\mathbf{N})$, where $\mathbf{N}$ is the adjacency matrix and $\rho \in (0, 1)$ is a predetermined correlation coefficient. It is possible to estimate ρ as a model parameter, but doing so requires an eigendecomposition of the Moran's operator ($\mathcal{O}(m^3)$) at each iteration of the MCMC algorithm, which can negate the computational gains of the PICAR approach. Another alternative is setting $\mathbf{Q} = I$, where the mesh nodes $\tilde{\mathbf{W}}$ and re-parameterized spatial random effects δ are uncorrelated.

3.3.4 Computational Gains

The PICAR approach requires shorter computational times per iteration as well as fewer iterations for the Markov chain to converge. The computational speedup results from bypassing expensive matrix operations (e.g. Cholesky decomposition) and by decorrelating and reducing the dimensions of the spatial random effects. The computational cost is dominated by the matrix-vector multiplication $\mathbf{AM}\delta$, where $\mathbf{AM}$ is the $n \times p$ basis function matrix constructed prior to model fitting and δ are reparameterized spatial random effects (basis coefficients). The PICAR approach has a computational complexity of $\mathcal{O}(np)$ as opposed to $\mathcal{O}(n^3)$ for the full hierarchical spatial model. Figure 3.5 illustrates the computational speedup offered by the PICAR approach. As we increase the dimensionality of the observations n, the full hierarchical model quickly becomes computationally prohibitive. On the other hand, we can model the data using PICAR approach within the order of minutes. The computation times are based on a single 2.2 GHz Intel Xeon E5-2650v4 processor. All the code was run on the Pennsylvania State University Institute for CyberScience-Advanced CyberInfrastructure (ICS-ACI) high-performance computing infrastructure.

We examine mixing in MCMC algorithms within the context of spatial generalized linear mixed models (SGLMMs). Here, the PICAR approach generates a faster mixing MCMC algorithm than the re-parameterization method (Rep-SGLMM) (Christensen et al., 2006), an approach designed to improve mixing for SGLMMs. This is corroborated by the larger effective sample size per second (ES/sec), the rate at which independent samples are generated by the MCMC algorithm. Larger values of ES/sec indicates faster mixing. In the binary simulated example (Section 3.4.1), the effective samples per second value for the PICAR approach is roughly 345 times larger than the Rep-SGLMM approach. Additional details on the comparative study are provided in Section 3.4.1.

We also show how to implement the PICAR approach in the `stan` programming language (Carpenter et al., 2017). `stan` provides a full Bayesian statistical inference via the No U-Turn Sampler (Hoffman and Gelman, 2014) and Hamiltonian Monte Carlo (Neal, 2011). Once the model is specified, `stan` automatically generates a fast mixing MCMC algorithm. `stan`'s MCMC algorithm tends to be slower per iteration, but this balanced by the fast mixing (and low autocorrelation) of

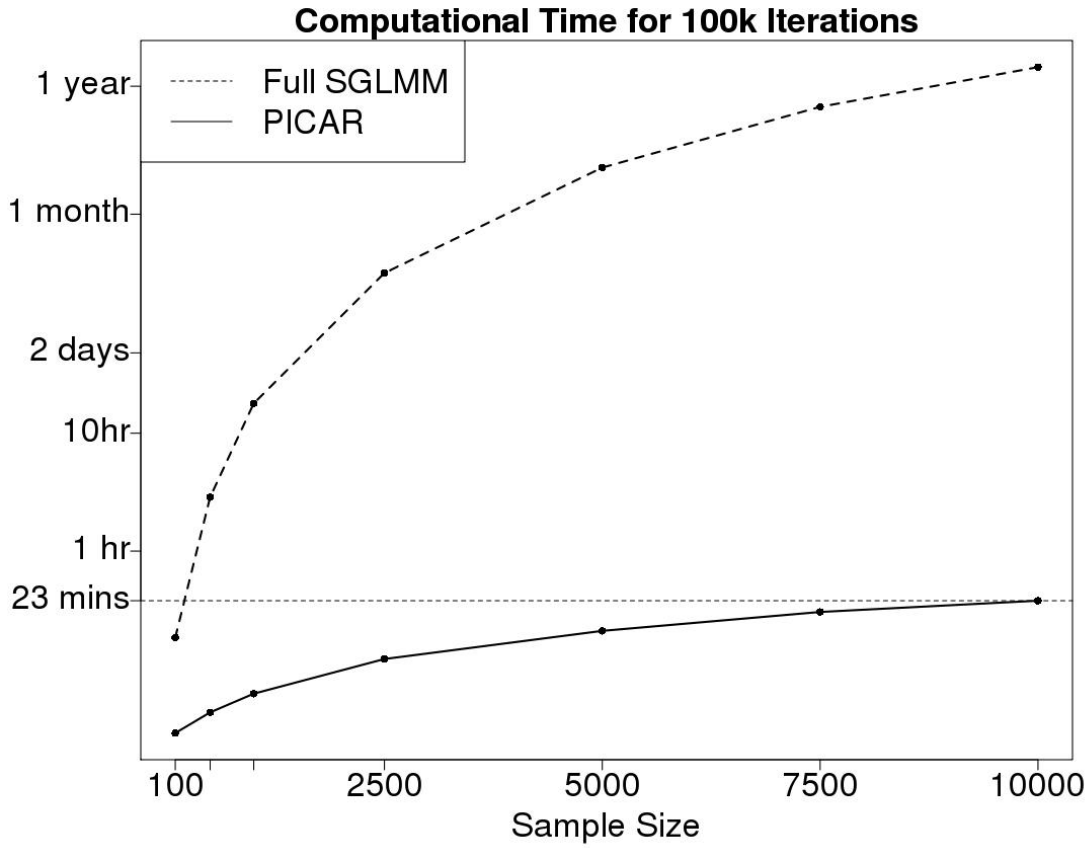

Figure 3.5. Computational time for 10^5 iterations versus sample size (n) for the full spatial generalized linear mixed model (SGLMM) and the PICAR approach with Moran's rank $p = 50$.

the resulting Markov chain. In Section 3.4.2, we use the R package **rStan** (Stan Development Team, 2019), in conjunction with the PICAR approach, to fit a spatially-varying coefficients model (Section 3.2.2).

For PICAR, the two major computational bottlenecks are constructing the Moran's operator (Section 3.3.1) and computing its eigencomponents. The Moran's operator requires the matrix operation $(\mathbf{I} - \mathbf{1}\mathbf{1}'/m)\mathbf{W}(\mathbf{I} - \mathbf{1}\mathbf{1}'/m)$ and $2m^3 - m^2$ floating point operations (FLOPs), which may be computationally prohibitive for large datasets. We reduce computational costs by leveraging the embarrassingly parallel operations as well the sparsity of the weights matrix $\mathbf{N}$. We use the sparse matrix R package **Matrix** (Bates and Maechler, 2019) to reduce costs for the operation $\Sigma = (\mathbf{I} - \mathbf{1}\mathbf{1}'/m)\mathbf{N}$. Then, we partition the resulting matrix Σ into K mutually exclusive $\frac{1}{K} \times n$ sub-matrices Σ_k for $k = 1, ..., K$. By parallelizing across

K processors, we can quickly construct the partial Moran's Operator $MO_k = \Sigma_k(I - 11'/m)$ for $k = 1, ..., K$. Finally, we generate the full Moran's Operator by combining the MO_k's as so $MO = \begin{bmatrix} MO_1 \\ \vdots \\ MO_K \end{bmatrix}$.

We can compute the k eigencomponents of the Moran's Operator using a partial eigendecomposition approach such as the Implicitly Restarted Arnoldi Method (Lehoucq et al., 1998) from **RSpectra** package (Qiu and Mei, 2019). Since the PICAR approach typically selects a rank($\mathbf{M}$) $\ll n$, there is no need to perform a full eigendecomposition of the Moran's Operator $\mathbf{M}$.

3.4 Simulation Study

We demonstrate the application of PICAR to several hierarchical spatial models:

1. **Binary data:** We chose this example because it is computationally feasible to fit the full spatial model ("gold standard") for comparison. We find that our approach is computationally efficient and provides comparable results to the gold standard.

2. **Poisson data with spatially varying coefficients:** With this example we show how PICAR can be easily extended to user-specified models, for instance by using simple code in the probabilistic programming language `stan`.

3. **Ordered categorical data:** We show how our approach efficiently fits a model for which there is no publicly available software.

For each data type, we generate 100 spatial datasets with locations on the unit domain $[0, 1]^2$. Each dataset consists of $1,400$ locations randomly chosen on the spatial domain. We use $1,000$ observations to fit the hierarchical spatial model and reserve 400 observations for validation. We chose $n = 1,000$ so that we could compare against a gold standard method (below) for which higher dimensions can be computationally prohibitive. We compare PICAR's performance across varying ranks of the Moran's operator $p = \{10, 50, 75, 100, 200\}$ as well as the

three different precision matrices (Independent, ICAR, and CAR) introduced in Section 3.3.3. The CAR precision matrix uses a correlation parameter $\rho = 0.5$. In the second comparison (precision matrices), we use the automated heuristic (Section 3.3.3) to select the rank. We compare PICAR's inference and prediction performance against a gold standard approach that uses MCMC based on the reparameterization in Christensen et al. (2006) which provides improved mixing over default MCMC algorithms.

3.4.1 Binary Data

All 100 datasets share common regression coefficients $\beta = (1, 1)^T$, and the random effects ($w(s)$'s) are generated using the Matérn covariance function with $\nu = 2.5$, $\sigma^2 = 1$, and $\phi = 0.2$. The covariance function has the form (cf. Rasmussen and Williams, 2006)

$$C(h) = \sigma^2\Big(1 + \frac{\sqrt{5}|h|}{\phi} + \frac{5|h|^2}{3\phi^2}\Big)\exp\Big(-\frac{\sqrt{5}|h|}{\phi}\Big),$$

where $|h|$ is the Euclidean distance between locations.

For the PICAR approach, the triangular mesh consists of $m = 1,649$ vertices. We use a vague multivariate normal prior for the regression coefficients $\beta \sim N(\mathbf{0}, 100I)$. As in Hughes and Haran (2013), we chose a gamma prior for $\tau \sim G(0.5, 2000)$. Here, we follow the prior belief that the covariates can sufficiently explain the data because large values for τ implies small variances for the random effects. For the binary and count data, we use Gibbs updates for τ and random-walk Metropolis-Hastings updates for β with proposal $\beta^{(i+1)} \sim N(\beta^{(i)}, \hat{\mathbf{V}})$, where $\hat{V}$ is the asymptotic covariance matrix from fitting the classical generalized linear model. Finally, we update the reparameterized random effects δ using an all-at-once Metropolis-Hastings update with a multivariate normal proposal centered at the parameters of the previous iterations as in Guan and Haran (2018). For the PICAR approach, we ran $300,000$ iterations of the MCMC algorithm.

We model spatial binary observations generated via the logit-link function $\text{logit}(p) = \log\{\frac{p}{1-p}\}$. We select one sample (from the 100 generated samples) as the dataset for the comparative analysis. When comparing across ranks, we

use the precision matrix from the ICAR model $Q = (\mathbf{W1} - \mathbf{W})$. We examine the out-of-sample cross-validated mean squared prediction error (CVMSPE):

$$\text{CVMSPE} = \frac{1}{n_{CV}} \sum_{i=1}^{n_{CV}} (Y_i^* - \hat{Y}_i^*)^2,$$

where $n_{CV} = 400$, Y_i^*'s denote the i-th value in the validation sample, and $\hat{Y}_i^*$'s are the predicted values at the i-th location.

Choice of Rank and Precision Matrix

Table 3.1 presents the parameter estimates, prediction results, and computational times for each rank p of the Moran's Operator. Results suggest that the rank is a key driver for predictive performance and parameter estimation. The PICAR approach is not sensitive to the chosen precision matrix $\mathbf{Q}$, as the results are similar across precision matrices (Table 3.2). The PICAR approach improves mixing in the MCMC algorithm as shown by the larger effective samples per second (ES/sec) compared to the gold standard approach. For model parameters β_1 and β_2, PICAR yields an ES/sec of 29.4 and 40.2 respectively and the gold standard returns an ES/sec 0.19 and 0.29 respectively. For the random effects $\mathbf{W}$, the average ES/sec is 5.8 for the PICAR approach and 0.016 for the gold standard, an improvement by a factor of roughly 345.

Rank	β_1 (95% CI)	β_2 (95% CI)	CVMPSE	Time (min)
10	1.04 (0.77,1.31)	0.91 (0.64,1.16)	0.3	9.73
22	**1.09 (0.82,1.37)**	**0.93 (0.67,1.2)**	**0.27**	**10.73**
50	1.12 (0.83,1.41)	0.95 (0.67,1.23)	0.28	11.14
75	1.14 (0.85,1.44)	0.98 (0.69,1.26)	0.28	11.62
100	1.2 (0.9,1.5)	1 (0.71,1.29)	0.29	12.28
200	1.34 (1.01,1.66)	0.99 (0.69,1.31)	0.32	15.13
Gold Standard	1.03 (0.77,1.3)	0.89 (0.63,1.16)	0.29	3624.43

Table 3.1. Simulated example with binary spatial observations. Parameter estimation, prediction, and model fitting time results across Moran's basis ranks. Bold font denotes the rank chosen by the automated heuristic.

Note that the PICAR approach is computationally efficient, and it also outperforms the gold standard approach in prediction. This is consistent with results from another basis representation approach, the latent conjugate model (Bradley

Precision Matrix	β_1 (95% CI)	β_2 (95% CI)	CVMPSE	Time (min)
Ind	1.07 (0.8,1.34)	0.92 (0.65,1.18)	0.28	9.53
ICAR	1.09 (0.82,1.37)	0.93 (0.67,1.2)	0.27	10.73
CAR	1.05 (0.79,1.33)	0.91 (0.65,1.18)	0.27	10.38
Gold Standard	1.03 (0.77,1.3)	0.89 (0.63,1.16)	0.29	3624.43

Table 3.2. Simulated example with binary spatial observations. Parameter estimation, prediction, and model fitting time results across precision matrices.

et al., 2019), which also outperforms the full SGLMM in computational cost and predictive ability. This may be attributed to the flexibility of PICAR's basis representation of the latent spatial field.

Simulation Study

We examine boxplots for the parameter estimates of β_1 and β_2 across the 100 samples (Supplement). The point estimates from the PICAR approach are distributed narrowly around the true values. The distribution of the point estimates remain similar across the choice of precision matrix $\mathbf{Q}$. The coverage proportions (0.89 for β_1, 0.91 for β_2) are close to but lower than the nominal coverage value (0.95).

3.4.2 Poisson Data with Spatially Varying Coefficients

In this section, we incorporate the PICAR approach to the spatially varying coefficients model using the `stan` programming language (Carpenter et al., 2017), a popular computing framework for Bayesian inference. As in the binary example, we generate $n = 1,400$ observations using the specified model parameters β, ϕ, σ^2. We assign one set of spatially varying coefficients $\beta_1(s)$ corresponding to the first covariate X_1. $\mathbf{W} = (w(s_1), ..., w(s_n))$ are the spatial random effects and $\mathbf{B} = (\beta_1(s_1), ..., \beta_1(s_n))$ is the n-dimensional vector of the spatially varying coefficients for each location $s_i \in \mathcal{D}$. Here, $(\mathbf{W}, \mathbf{B})^T \sim \mathcal{N}(0, \mathcal{N}(\mathbf{0}, R_\phi \otimes \mathbf{T}))$, where R_ϕ is the correlation function from the binary case and $\mathbf{T} = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 0.2 \end{bmatrix}$. In the PICAR framework, we approximate the spatial processes $\mathbf{B}$ and $\mathbf{W}$ as $\mathbf{B} \approx [\mathbf{AM}\delta_\beta](s)$ and $\mathbf{W} \approx [\mathbf{AM}\delta_\mathbf{w}](s)$, where $\mathbf{A}$ is the $n \times m$ projector matrix, $\mathbf{M}$ is the $m \times p$ Moran's basis function matrix, and δ_β and δ_w are the corresponding basis coefficients. Note

that we are modeling both spatial process $\mathbf{B}$ and $\mathbf{W}$ as independent spatial processes with no cross-correlations. Extensions to multivariate models are promising avenues for future research. The PICAR specification of this hierarchical model is as follows:

$$
\begin{aligned}
&\textbf{Data Model:} && Z(s)|\eta(s) \sim f(\eta(s)),\\
& && \eta(s) = X(s)\beta + X_1(s)[\mathbf{AM}\delta_\beta](s) + [\mathbf{AM}\delta_{\mathbf{w}}](s),\\
&\textbf{Process Model:} && \delta_\beta \sim \mathcal{N}(0, \tau_\beta^{-1}(\mathbf{M}'\mathbf{Q}_\beta\mathbf{M})^{-1}),\\
& && \delta_w \sim \mathcal{N}(0, \tau_w^{-1}(\mathbf{M}'\mathbf{Q}_{\mathbf{w}}\mathbf{M})^{-1}),\\
&\textbf{Parameter Model:} && \tau_\beta \sim G(\alpha_{\tau 1}, \beta_{\tau 1}), \quad \tau_w \sim G(\alpha_{\tau 2}, \beta_{\tau 2}), \quad \beta \sim N(\mu_\beta, \Sigma_\beta),
\end{aligned}
$$

where X_1 is the first column of the $n \times 2$ design matrix $\mathbf{X}$. $\mathbf{Q}_\beta$ and $\mathbf{Q}_{\mathbf{w}}$ are the $m \times m$ precision matrix for the mesh vertices and τ_β and τ_w are the precision parameters. $\alpha_{\tau 1}, \beta_{\tau 1}, \alpha_{\tau 2}, \beta_{\tau 2}, \mu_\beta$, and Σ_β are the hyperparameters. Note that the $p-$dimensional vectors δ_β and δ_w replace the n-dimensional vectors $\beta(s)$ and $w(s)$ in the traditional spatially varying coefficients model (Section 3.2.2).

Results

We compare PICAR's performance across varying ranks for the Moran's operator. Similar to the binary case, the chosen rank of the Moran's operator drives predictive performance and parameter estimation (Table 3.3). Here, we achieve a low CVMSPE using the rank selected via the automated heuristic ($p = 63$). Model fitting times increase with respect to the chosen rank of the Moran's operator. Using `stan`, we obtained an effective sample size of $\sim 5,000$ for all parameters, random effects $w(s)$, and the spatially varying coefficients $\beta(s)$.

3.4.3 Ordered Categorical Data

We use the PICAR approach to model ordered categorical data. For the simulation study, we generate 100 samples of ordered categorical spatial observations using the spatial cumulative-logit model from Section 3.2.2. Similar to the binary case, we select one sample to be the focus of our comparative analysis. The true cut-off parameters (Section 3.2.2) are $\theta_1 = 0$, $\theta_2 = 1$ and $\theta_2 = 2$. To avoid identifiability

Rank	β_1 (95% CI)	β_2 (95% CI)	CVMPSE	Time (min)
10	0.92 (0.81,1.02)	0.94 (0.85,1.03)	3.59	0.52
50	0.77 (0.54,1.01)	0.99 (0.89,1.09)	2.30	7.68
63	**0.77 (0.48,1.06)**	**1.02 (0.9,1.12)**	**2.31**	**13.15**
75	0.86 (0.55,1.16)	1.03 (0.93,1.14)	2.68	22.07
100	0.95 (0.55,1.33)	1.06 (0.94,1.17)	2.82	46.55
200	1.07 (0.7,1.49)	1.07 (0.95,1.19)	3.80	226.49

Table 3.3. Simulated example with spatially varying coefficients. Model fit using `stan` programming language. Parameter estimation, prediction, and model fitting time results across Moran's basis ranks. Bold font denotes the rank chosen by the automated heuristic.

issues in model fitting, we fix the first cutoff $\theta_1 = 0$. We also reparameterize the cut-off parameters into α_1 and α_2 as described in Section 3.2.2. To assess predictive performance, we examine the out-of-sample misprediction rate (MPR), or the proportion of incorrect predictions, and the loss function is:

$$\text{MPR} = \frac{1}{n_{CV}} \sum_{i=1}^{n_{CV}} I_{(Y_i^* \neq \hat{Y}_i^*)},$$

where $\hat{Y}^*$ are the predicted values and Y^* are the true values at the validation locations.

The automated heuristic chose a rank of $p = 23$, which yields comparable parameter estimation and predictive ability to the gold standard (see Supplement for details). While predictive ability does not vary considerably across rank, the chosen rank is important for parameter estimation. The PICAR approach is not sensitive to the chosen precision matrix Q, as the inferential and predictive performances do not vary across precision matrices. Similar to the binary case, we observe faster mixing with the PICAR approach compared to the gold standard. For model parameters β_1 and β_2, the PICAR approach exhibits an ES/sec of 30.7 and 30.4 respectively and the gold standard yields an ES/sec 0.034 and 0.034 respectively. For the random effects, the PICAR approach has an average ES/sec of 4.4 and 0.002 for the gold standard, an improvement by a factor of approximately $1,487$.

For the simulation study, we examine boxplots for the parameter estimates for β_1, β_2, α_1, and α_2 across all 100 samples. In Figure 3.6, we see that the parameter

estimates are centered around the true parameter values. Similar to the binary case, our coverage proportions are very close (0.91 for β_1, 0.92 for β_2, 0.93 for α_1, 0.88 for α_2), but slightly lower than the nominal coverage (0.95).

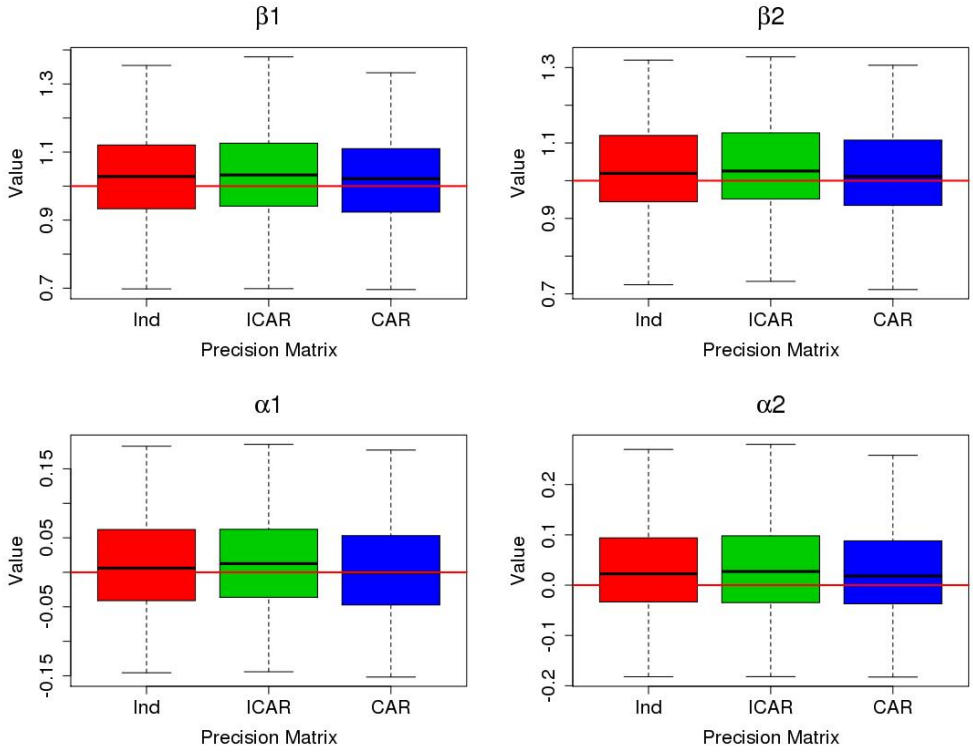

Figure 3.6. Ordinal data simulation study: distribution of posterior mean estimates for parameters β_1 (top left) β_2 (top right), α_1 (bottom left), and α_2 (bottom right) for three different precision matrices - Independent (red), ICAR (green), and CAR with $\phi = 0.5$ (blue). The red horizontal line denotes the true parameter values. The automated heuristic selects the appropriate rank p of the Moran's operator $\mathbf{M}$. Note that the default precision matrix for the PICAR approach is the ICAR precision matrix (green). Distributions are similar across precision matrices.

3.5 Real Data Examples

Through the following applications, we demonstrate how the PICAR approach readily scales up to higher-dimensional spatial datasets:

1. **Parasitic infestation of dwarf mistletoe:** We model $n = 22,888$ binary spatial observations in around 4 hours, while this is computationally

prohibitive under the traditional hierarchical modeling framework.

2. **Volunteer-driven watershed quality surveys:** We fit a hierarchical spatial model to $n = 5,561$ ordered categorical spatial observations within 35 minutes. Fitting the full model is too expensive, and there are no publicly available approaches for fitting this model efficiently.

3.5.1 Binary Data: Parasitic Infestation of Dwarf Mistletoe

In Minnesota, the eastern spruce dwarf mistletoe (Arceuthobium pusillum) are a parasitic species that affect the longevity and quality of its host, the black spruce (Picea mariana) (Geils and Hawksworth, 2002). This infestation has economic ramifications because the black spruce are valuable resources used to produce high quality paper. We use a dataset from Hanks et al. (2011) originally obtained from the Minnesota Department of Natural Resources (DNR) forest inventory. The response is a binary incidence of dwarf mistletoe at $n = 25,431$ black spruce stands (i.e. location hosting the black spruce samples). We randomly sample $22,888$ observations to fit our model and reserve $2,543$ observations for validation. Covariates include the: (1) average age of trees in the stand (2) basal area per acre of trees in the stand; (3) average canopy height; and (4) volume of the stand in cords, a unit of measurement. We fit a hierarchical spatial model, specifically the SGLMM with a logit link function, using the PICAR approach. We construct a triangular mesh with $m = 32,611$ mesh vertices, and the automated heuristic (Section 3.3.3) chose a rank of $p = 520$ for the Moran's basis functions matrix.

The PICAR approach required around 4 hours to fit the model. Specifically, it took 2 hours to run 10^5 iterations of the MCMC algorithm, 10 minutes to generate the Moran's operator (Section 3.3.1) via parallel computing across 100 processors, and 1.7 hours to calculate the first $1,000$ eigencomponents using the **Spectra** C++ library. Comparison with the full SGLMM is infeasible as fitting the full SGLMM (Section 3.2.1) to this dataset is computationally prohibitive. The posterior predictive map displays similar spatial patterns between the predicted and true values for the validation sample (Figure 3.7).

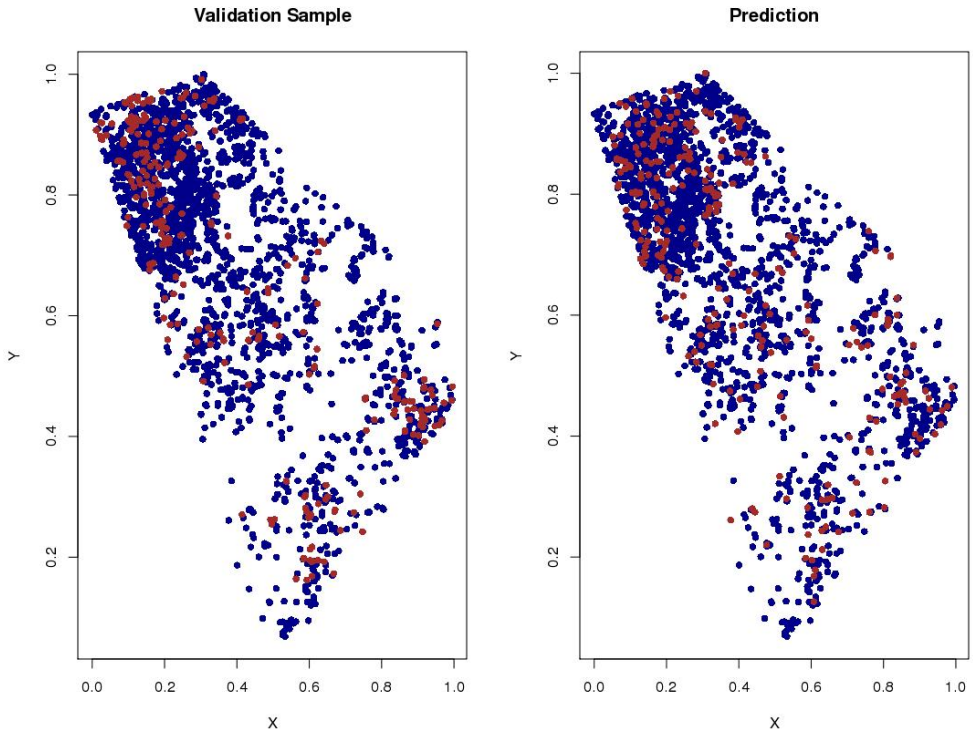

Figure 3.7. Observed (left) and predicted (right) dwarf mistletoe presence and absence at the validation sample locations. Red points denote the presence of dwarf mistletoe and blue points denote absence.

3.5.2 Ordered Categorical Data: MD Stream Waders

Beginning as a pilot program in 2000, the Maryland Stream Waders (MSW) program is a statewide volunteer stream monitoring program managed by the Maryland Department of Natural Resources' (DNR) and the Monitoring and Non-Tidal Assessment Division (MANTA). The MSW program was designed to supplement the data from the Maryland Biological Stream Survey (MBSS) by increasing the density of sampling sites for use in stream and watershed assessments. To illustrate, the MBSS samples are collected at the watershed level (around 70 square miles), while the MSW samples at a smaller scale subwatersheds level (around 8 square miles each). For the samples collected at each site, the DNR laboratory calculated an Benthic Index of Biotic Integrity (BIBI) was calculated (on a 1 to 5 scale). Each site was rated either Good (BIBI 4-5), Fair (BIBI 3-3.9), or Poor (BIBI 1-2.9) (Stribling et al., 1998).

A total of $6,951$ samples were collected within a 17-year time period (2000-2017) at irregular sampling locations (Maryland's Mapping and GIS Data Portal, 2018). We fit the model using $5,561$ randomly selected observations and validate the model with the remaining $1,390$ samples. We model the observations using the spatial cumulative-logit models (Section 3.2.2) with just an intercept term. We generated a mesh with $m = 8,810$ nodes and the automated heuristic chose a rank of $p = 653$. The time to fit the spatial cumulative-logit model via PICAR is around 35 minutes. We estimate fitting the full hierarchical spatial model would require months to provide similarly accurate inference.

3.6 Discussion

In this study, we propose a fast extendable projection-based approach (PICAR) for modeling a wide range of hierarchical spatial models. In cases where it is possible to fit the full hierarchical spatial model, we show that our approach yields comparable results in terms of both inference and prediction. We also provide a variety of other examples that illustrate the flexibility of the PICAR approach as well as the ease with which non-experts can specify and efficiently fit their own hierarchical spatial models. We show that our approach is computationally efficient, scales up to higher dimensions, automated, and extendable to a variety of hierarchical spatial models. We provide an example of a hierarchical spatial model (ordinal spatial data) that cannot be fit using existing publicly available code but can be easily fit using PICAR. Moreover, we show that our approach is amenable to implementation in a programming language for Bayesian inference (`stan`). As shown in our real-data applications, our approach scales well to higher dimensions. Where other approaches may be computationally infeasible, we can fit a high-dimensional hierarchical spatial model within hours.

The computational complexity for the PICAR approach is driven by matrix-vector multiplications, which can be readily parallelized. With efficient parallelization methods, we expect our approach to scale up to hundreds of thousands of data points. Even though an eigendecomposition is only carried out once in our approach, methods such as Nyström method (Williams and Seeger, 2001) or random projections (Banerjee et al., 2013; Guan and Haran, 2018) can further reduce

costs via an approximate eigendecomposition of the Moran's operator. There may be other methods to improve our automated heuristic for rank selection such as implementing a screening process for the relevant basis functions via a variable selection approach like LASSO (Tibshirani, 1994). Extending the PICAR approach to spatio-temporal or multivariate spatial processes as well as computer model calibration with non-Gaussian model outputs may provide fruitful avenues for future research.

3.7 Acknowledgments

We are grateful to Ephraim Hanks, Erin Schliep, Yawen Guan, Jaewoo Park, and Klaus Keller for helpful discussion. This work was partially supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Earth and Environmental Systems Modeling, MultiSector Dynamics, Contract No. DE-SC0016162 and by the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507. This study was also co-supported by the Penn State Center for Climate Risk Management. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Department of Energy, the National Science Foundation, or other funding entities. Any errors and opinions are, of course, those of the authors. We are not aware of any real or perceived conflicts of interest for any authors.

Chapter 4

Modeling and Computation for High-dimensional Zero-Inflated Spatial Data

In this section, I discuss modeling approaches for zero-inflated spatial data and propose a computationally efficient method for fitting high-dimensional zero-inflated spatial models. First, I introduce the general framework for spatial two-part models, a popular model for zero-inflated spatial observations. Next, I propose a computationally efficient approach for fitting high-dimensional two-part models using the PICAR approach (Chapter 3). Finally, I demonstrate the proposed approach on multiple simulated examples as well as a high-dimensional species abundance dataset. This chapter is currently in preparation for submission.

4.1 Introduction

Zero-inflated spatial data are spatially dependent observations characterized by an excess of zeros. Observations can be discrete counts or semi-continuous, where the non-zero values are positive real numbers. Zero-inflated spatial data are common in many disciplines; for example, counts of harbor seals on glacial ice (Hoef and Jansen, 2007), annual mental health expenditures among US federal employees Neelon et al. (2011), and the number of torrential rainfall events in a region of

interest Lee and Kim (2017). Standard probability distributions are not sensible for modeling zero-inflated data (cf. Agarwal et al., 2002; Rathbun and Fei, 2006; Lambert, 1992a) as they are unable to account for the large proportion of zeros. Moreover, poor model choice may lead to over- or under-dispersion, where the observed variance is higher or lower, respectively, than the variance of the fitted model.

Two-part models (Mullahy, 1986; Lambert, 1992b) are typically used to model zero-inflated spatial observations (Agarwal et al., 2002; Hoef and Jansen, 2007; Olsen and Schafer, 2001, .cf). Two-part models consist of two processes, the occurrence and prevalence processes. The occurrence process dictates whether a zero or non-zero value is observed at a particular location, and the prevalence process determines the value of non-zero observations (exceptions may apply). The occurrence and prevalence processes are typically modeled using spatial generalized linear mixed models (SGLMMs). For high-dimensional data, fitting SGLMMs can be computationally burdensome due to arge matrix operations and slow mixing Markov Chain Monte Carlo (MCMC) algorithms.

Past studies propose novel modeling approaches for zero-inflated spatial data, but these methods may not scale well to larger datasets. Two-part models have been fit using Guass-Hermite quadrature (Min and Agresti, 2005), expectation-maximization (Lambert, 1992b; Roeder et al., 1999), restricted maximum quasi-likelihoods (Kim et al., 2012), or Monte Carlo maximum likelihood methods (Lyashevska et al., 2016). However, such approximations do not scale well with high-dimensional spatial random effects. In the Bayesian framework, the literature focuses on improving the sophistication of zero-inflated spatial models. Studies have modeled spatio-temporal dependence (Fernandes et al., 2009; Neelon et al., 2016a; Arcuti et al., 2016), addressed overdispersion (Gschlößl and Czado, 2008; Lee et al., 2016), used skewed distributions (Dreassi et al., 2014; Liu et al., 2016), used t-distributions to model heavy tailed behavior (Neelon et al., 2015), and modeled prevalence with scale mixtures of normal distributions (Fruhwirth-Schnatter and Pyne, 2010) or Student-t processes (Bopp et al., 2020). However, there is a dearth of research on addressing computational issues with large zero-inflated spatial datasets, namely the high-dimensional correlated spatial random effects and slow mixing of MCMC algorithms. Wang et al. (2014) models the presence and

abundance of Atlantic cod in 1325 locations along the Gulf of Maine using predictive processes (Banerjee et al., 2008). Another study facilitates posterior sampling for zero-inflated negative binomial distributions (ZINB) by using latent variables that are represented as scale mixtures of normal distributions (Neelon et al., 2018).

In this study, we introduce a computationally efficient approach for fitting a broad range of two-part models to high-dimensional zero-inflated spatial observations. We use the projection-based intrinsic conditional autoregression (PICAR) approach (Lee and Haran, 2019) to reducing the dimensions of and correlation between the spatial random effects in two-part spatial models. The PICAR method represents the spatial random effects with empirical basis functions. Various basis representations have been directly or indirectly used to model spatial data, for instance in the predictive process approach (Banerjee et al., 2008), random projections (Guan and Haran, 2018, 2019; Banerjee et al., 2013; Park and Haran, 2019), Moran's basis for areal models (Hughes and Haran, 2013), stochastic partial differential equations (Lindgren et al., 2011), kernel convolutions (Higdon, 1998), eigenvector spatial filtering (Griffith, 2003), and multi-resolution basis functions (Nychka et al., 2015; Katzfuss, 2017), among others. We utilize a non-parametric set of basis functions based on the Moran's I statistic and piece-wise linear basis functions. To our knowledge, this is the first approach that readily lends itself to user-specified spatial two-part models for zero-inflated spatial data while also reducing computational costs for large datasets. We demonstrate the applicability of PICAR via simulation studies as well as an species abundance dataset of benthic invertebrates.

In section 4.2, we introduce the two-part modeling framework for zero-inflated data. In section 4.3, we provide an overview of spatial two-part models and examine the associated modeling and computational challenges. Then, in section 4.4, we propose a computationally efficient approach to fit high-dimensional spatial two-part models. We also outline the tuning mechanisms and computational advantages of our approach. We demonstrate the proposed approach using four simulated examples generated from popular spatial two-part models (section 4.5) as well as a high-dimensional ecological dataset (section 4.6). Finally, a brief summary and directions for future research are provided in Section 4.7.

4.2 Two-part Models For Zero-inflated Data

In this section, we introduce the "two-part" modeling framework for zero-inflated data. *Two-part models* (Mullahy, 1986; Lambert, 1992b) are a popular class of models for modeling zero-inflated data. These models are comprised of two random variables: (1) the occurrence random variable O, which specifies the structural zero and non-zeros cases; and (2) the prevalence random variable P that generates the positive values for the structural non-zero cases. For special cases, the prevalence random variable P can also generate zeros. In two-part models, the zero-inflated observation Z is generated as follows:

$$Z = \begin{cases} 0 & \text{if } O = 0 \\ P & \text{if } O = 1. \end{cases}, \tag{4.1}$$

where O and P are the latent occurrence and prevalence random variables, respectively. In the univariate case, the occurrence variable $O \in \{0, 1\}$ is modeled as a Bernoulli random variable with probability $\pi \in (0, 1)$ (i.e., $O \sim \text{Bern}(\pi)$). The prevalence variable is distributed as $P \sim f(\theta)$ where $f(\theta)$ is a discrete probability mass function or continuous probability density function with prevalence model parameter θ.

Two-part models for zero-inflated data typically fall into two classes:

1. **Hurdle Models:** The occurrence random variable O solely specifies the zero-valued observations. The prevalence random variable P generates positive values for the non-zero-valued observations. In the discrete case, $f(\cdot)$ is a zero-truncated distribution such as the zero-truncated Poisson or the zero-truncated negative binomial distribution. For semi-continuous observations, $f(\cdot)$ is a probability density function with positive support such as a log-normal or gamma distribution.

2. **Mixture Models:** Both the occurrence O and prevalence random variables P specify the zero-valued observations. The occurrence random variable identifies the structural zero-valued observations. The prevalence random variable P generates both zeros and positive values for structural non-zero-valued observations. In the discrete case, $f(\cdot)$ is a non-degenerate distribution

such as the Poisson or Negative-Binomial distribution. For semi-continuous observations, $f(\cdot)$ can be a censored model such as a Tobit Type I.

For the univariate case, the likelihood function $\tilde{f}_Z\big(z; O, P\big)$ for two-part models is as follows:

$$\tilde{f}_Z\big(z; O, P\big) = \begin{cases} \pi + (1-\pi) \times f(0;\theta), & \text{if } z = 0 \\ (1-\pi) \times f(z;\theta), & \text{if } z > 0. \end{cases}, \tag{4.2}$$

For two part models, the expectation is defined as $E[Z|\pi,\theta] = \pi E_f[Z|\theta]$ and the variance is $Var[Z|\pi,\theta] = \pi(1-\pi)E_f[Z|\theta]^2 + \pi Var_f[Z|\theta]$ where $E_f[Z|\theta]$ and $Var_f[Z|\theta]$ denote the expectation and variance of a random variable with probability distribution $f(\cdot|\theta)$.

Alternative distributions for $f(\cdot)$ can result in richer and more flexible two-part models. For count data, examples include the Poisson, negative binomial, zero-truncated Poisson (Lambert, 1992a), translated Poisson (Hoef and Jansen, 2007), zero-truncated negative binomial (Mwalili et al., 2008), generalized Poisson (Gschlößl and Czado, 2008), and binomial distributions (Hall, 2000). In the semi-continuous case, the lognormal distribution may not be appropriate due to the lack of symmetry or fatter tails exhibited by the observations. Past studies have used skewed distributions (Dreassi et al., 2014; Liu et al., 2016), t-distributions to model heavy tailed behavior (Neelon et al., 2015), or modeled the prevalence process using scale mixtures of normal distributions (Fruhwirth-Schnatter and Pyne, 2010).

For independently and identically distributed observations $\mathbf{Z} = \{Z_1, ..., Z_n\}$, statistical inference consists of estimating model parameters π and θ by maximizing the joint likelihood function $\tilde{f}_{\mathbf{Z}}\big(\mathbf{z}; O, P\big) = \prod_{i=1}^{n} \tilde{f}_Z\big(z_i; O, P\big)$. Bayesian hierarchical models are well-suited for two-part models because two-part models have a natural hierarchical structure, as shown by the data-generating model (Equation 4.1) and latent random variables (O and P). The Bayesian hierarchical framework

for two-part models is as follows:

Data Model:	$Z\|O, P \sim \tilde{f}_Z\big(z; O, P\big)$
Process Model:	$O\|\pi \sim \text{Bern}(\pi)$
	$P\|\theta \sim f(\theta)$
Parameter Model:	Priors for π and θ

4.3 Spatial Two-part Models

In this section, we discuss the two-part modeling framework for zero-inflated spatial data. Then, we present an overview of the modeling and computational challenges associated with fitting these models. Past research have extended two-part models to spatially dependent zero-inflated observations (Hoef and Jansen, 2007; Neelon et al., 2016a; Wang et al., 2014, cf.). Spatial two-part models have a similar modeling framework to the univariate case in Section 4.2; however, we allow the occurrence O and prevalence P random variables to vary in space. Here, we model occurrence and prevalence as spatial random processes $O(s)$ and $P(s)$.

Let $Z(s)$ be a zero-inflated observation for spatial location $s \subset \mathcal{D}$ within the spatial domain $\mathcal{D} \in \mathbb{R}^2$. The observation $Z(s)$ are generated as follows:

$$Z(s) = \begin{cases} 0 & \text{if } O(s) = 0 \\ P(s) & \text{if } O(s) = 1. \end{cases}, \tag{4.3}$$

where $O(s)$ and $P(s)$ are the spatial occurrence and prevalence processes, respectively. The occurrence process is specified as $O(s) \sim Bern(\pi(s))$ with spatially varying probabilities $\pi(s) \in (0, 1)$. The prevalence process is modeled as $P(s) \sim f(\theta(s))$ where $f(\theta(s))$ is a discrete or continuous probability distribution with spatially varying model parameters $\theta(s)$.

4.3.1 Spatial Generalized Linear Mixed Models

In the literature, particular focus has been placed on modeling the spatially-dependent occurrence $O(s)$ and prevalence $P(s)$ processes using spatial generalized linear mixed models (SGLMM)(cf. Agarwal et al., 2002; Rathbun and Fei, 2006;

Neelon et al., 2013; Recta et al., 2012). Spatial generalized linear mixed models are a popular choice for modeling non-Gaussian spatially dependent observations (Diggle et al., 1998).

Non-Gaussian spatial observations are typically modeled using spatial generalized linear mixed models (SGLMMs) (Diggle et al., 1998). Let $\{Z(s) : s \in \mathcal{D}\}$ be a non-Gaussian spatial random field. Assuming $Z(s)$ are conditionally independent given the latent random spatial field $\mathbf{W}$, the conditional mean $E[Z(s)|\beta, \mathbf{W}, \epsilon(s)]$ can be modeled through a linear predictor $\eta(s)$:

$$\eta(s) = g\{E[Z(s)|\beta, \mathbf{W}], \epsilon(s)\} = X(s)\beta + w(s) + \epsilon(s),$$

where $g(\cdot)$ is a known link function. Binary and count observations are two common types of non-Gaussian spatial data, and these can be modeled using the binary SGLMM with logit link and the Poisson SGLMM with log link, respectively. $X(s)$ is a set of k covariates associated with location s and β is a k-dimensional vector of coefficients. The micro-scale measurement errors or nugget are modeled as an uncorrelated Gaussian process with zero mean and variance τ^2 where $\epsilon(s) \sim N(0, \tau^2)$ for all $s \in \mathcal{D}$.

We impose spatial dependence by modeling the spatial random effects $\mathbf{W} = \{w(s) : s \in \mathcal{D}\}$ as a stationary zero-mean Gaussian process with a positive definite covariance function $C(\cdot)$. For a finite set of locations $s = (s_1, ..., s_n)$, the spatial random effects $\mathbf{W}$ are distributed as a multivariate normal distribution $\mathbf{W}|\Theta \sim N(0, C(\Theta))$ with covariance function parameters Θ and the covariance matrix $C(\Theta)$ where $C(\Theta)_{ij} = \text{cov}(w(s_i), w(s_j))$. The Matérn covariance function is a widely used class of stationary and isotropic covariance functions (Stein, 2012) with parameters $\Theta = (\sigma^2, \phi, \nu)$ such that:

$$C(s_i, s_j) = \sigma^2 \frac{1}{\Gamma(\nu)2^{\nu-1}} \Big(\sqrt{(}2\nu)\frac{h}{\phi}\Big)^{\nu} K_\nu\Big(\sqrt{(}2\nu)\frac{h}{\phi}\Big),$$

where $R(\phi)$ is the correlation matrix, $h = ||s_i - s_j||$ is the Euclidean distance between locations s_i and s_j, $\sigma^2 > 0$ is the partial sill or scale parameter of the process, and $\phi > 0$ is the range parameter for spatial dependence. $K_\nu(\cdot)$ is the modified Bessel function of the second kind where the smoothness parameter ν is commonly fixed prior to model fitting.

The Bayesian hierarchical framework for spatial generalized linear mixed models (SGLMMs) is:

Data Model:	$Z(s)\|\eta(s) \sim f(\eta(s))$
	$\eta(s) = g(\mathbb{E}[Z(s)\|\beta, \mathbf{W}), \epsilon(s)]) = X(s)\beta + w(s) + \epsilon(s)$
Process Model:	$\mathbf{W}\|\phi, \sigma^2 \sim N(0, \sigma^2 R_\phi)$
	$\epsilon(s)\|\tau^2 \sim N(\mathbf{0}, \tau^2)$
Parameter Model:	$\beta \sim p(\beta), \phi \sim p(\phi), \sigma^2 \sim p(\sigma^2), \tau^2 \sim p(\tau^2)$

4.3.2 Modeling Framework: HURDLE and Mixture Models

In this section, we outline the genreal modeling framework for spatial two-part models. Both processes, $O(s)$ and $P(s)$, are modeled as spatial generalized linear mixed models (SGLMMs) with the appropriate link functions. The occurrence process $O(s)$ is modeled as a Bernouilli random variable with either a probit or a logit link function. The linear predictor is defined as $\boldsymbol{\eta_o} = X\boldsymbol{\beta_o} + \mathbf{W_o} + \epsilon_o$, where $\mathbf{W_o} \sim \mathcal{N}(0, \sigma_o^2 R_{\phi_o})$ and $\epsilon_o \sim \mathcal{N}(0, \tau_o^2 \mathcal{I})$. Model fitting entails estimating the parameters $\beta_o, \phi_o, \sigma_o^2, \tau_o^2$ as well as the spatial random effects $\mathbf{W_o}$.

The prevalence process $P(s)$ follows a specific probability distribution based on the observation type (counts vs. semi-continuous) and structural assumptions. For HURDLE models, a zero-truncated distribution (e.g., zero-truncated Poisson, zero-truncated negative binomial, lognormal, or gamma) is a sensible choice for $f(\cdot)$. Mixture models utilize a distribution with non-negative support (e.g., Poisson, negative binomial, or Tobit model). Similar to the occurrence process, the prevalence process $P(s)$ is also modeled as an SGLMM with linear predictor $\boldsymbol{\eta_p} = X\boldsymbol{\beta_p} + \mathbf{W_p} + \epsilon_p$, where $\mathbf{W_p} \sim \mathcal{N}(0, \sigma_p^2 R_{\phi_p})$ and $\epsilon_p \sim \mathcal{N}(0, \tau_p^2 \mathcal{I})$. Here, the parameters $\beta_p, \phi_p \sigma_p^2, \tau_p^2$ and spatial random effects $\mathbf{W_p}$ must be estimated. To complete the Bayesian hierarchical framework, prior distributions are specified for the model parameters.

The Bayesian hierarchical framework for two-part models is as follows:

Data Model:	$\mathbf{Z}\|O(s), P(s) \sim \tilde{f}_Z\big(z; O(s), P(s)\big)$
Process Model:	$O(s)\|\pi(s) \sim \text{Bern}(\pi(s))$
	$P(s)\|\theta(s) \sim f(\theta(s))$
Sub-process Model 1:	$\pi(s)\|\eta_o(s) = g_o^{-1}(\eta_o(s))$
(Occurrence)	$\eta_o(s)\|\beta_o, W_o(s), \epsilon_o(s) = X(s)\beta_o + W_o(s) + \epsilon_o(s)$
	$\mathbf{W}_o = \{W_o(s_1), ..., W_o(s_n)\}$
	$\mathbf{W}_o\|\phi_o, \sigma_o^2 \sim \mathcal{N}(\mathbf{0}, \sigma_o^2 R_{\phi_o})$,
	$\epsilon(s)\|\tau_o^2 \sim \mathcal{N}(\mathbf{0}, \tau_o^2)$
Sub-process Model 2:	$\theta(s)\|\eta_p(s) = g_p^{-1}(\eta_p(s))$
(Prevalence)	$\eta_p(s)\|\beta_p, W_p(s), \epsilon_p(s) = X(s)\beta_p + W_p(s) + \epsilon_p(s)$
	$\mathbf{W}_p = \{W_p(s_1), ..., W_p(s_n)\}$
	$\mathbf{W}_p\|\phi_p, \sigma_p^2 \sim \mathcal{N}(\mathbf{0}, \sigma_p^2 R_{\phi_p})$
	$\epsilon(s)\|\tau_p^2 \sim \mathcal{N}(\mathbf{0}, \tau_p^2)$
Parameter Model:	Priors for $\beta_o, \phi_o, \sigma_o^2, \tau_o^2, \beta_p, \phi_p, \sigma_p^2,$ and τ_p^2

where $\tilde{f}_Z\big(z; O(s), P(s)\big)$ is the likelihood function of spatial two-part model. Based on Equation 4.3, $\tilde{f}_Z\big(z; O(s), P(s)\big)$ is defined as:

$$\tilde{f}_Z\big(z; O(s), P(s)\big) = \begin{cases} \pi(s) + (1 - \pi(s)) \times f(0; \theta(s)), & \text{if } z = 0 \\ (1 - \pi(s)) \times f(z; \theta(s)), & \text{if } z > 0. \end{cases}, \tag{4.4}$$

Similar to the univariate case, spatial two-part models typically fall into two classes - HURDLE and mixture models. The key difference between these two classes is the choice of $f(\cdot|\theta(s))$, the distribution used to model the prevalence process $P(s)$. Here, we describe four popular two-part spatial models.

HURDLE Model for Spatial Count Data

First, the HURDLE Poisson model is appropriate for zero-inflated count data where only the occurrence process $O(s)$ generates zeros. Here, the distribution for the prevalence process $P(s)$ is a zero-truncated distribution such as a zero-

truncated Poisson distribution (i.e. $f(\cdot|\theta(s)) = \frac{\theta(s)^z e^{-\theta(s)}}{z!(1-e^{-\theta(s)})}$). The associated likelihood function is:

$$\tilde{f}_Z\big(z; O(s), P(s)\big) = \begin{cases} \pi(s), & \text{if } z = 0 \\ (1-\pi(s)) \times \frac{\theta(s)^z e^{-\theta(s)}}{z!(1-e^{-\theta(s)})}, & \text{if } z > 0. \end{cases} . \tag{4.5}$$

Note that the prevalence process can be modeled using other zero-truncated discrete probability distributions such as the zero-truncated Negative binomial or a translated Poisson distribution (Hoef and Jansen, 2007).

HURDLE Model for Spatial Semi-continuous Data

The HURDLE lognormal model is appropriate for zero-inflated semi-continuous data where only the occurrence process $O(s)$ generates zeros. Here, the distribution for the prevalence process $P(s)$ is a multivariate lognormal distribution. The associated likelihood function is:

$$\tilde{f}_Z\big(z; O(s), P(s)\big) = \begin{cases} \pi(s), & \text{if } z = 0 \\ (1-\pi(s)) \times LN(z; \theta(s)), & \text{if } z > 0, \end{cases} . \tag{4.6}$$

where $LN(z;\theta(s)) = \frac{1}{z\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\ln z - \mu(s))^2}{2\tau^2}\right)$ is the probability density function of a lognormal distribution with parameters $\theta(s) = \{\mu(s), \tau^2\}$.

Mixture Model for Spatial Count Data

The zero-inflated Poisson (ZIP) model is popular choice for modeling zero-inflated count data where both the occurrence $O(s)$ and prevalence processes $P(s)$ generate zeros. The distribution for the prevalence process $P(s)$ is a Poisson distribution (i.e. $f(\cdot|\theta(s)) = \frac{\theta(s)^z e^{-\theta(s)}}{z!}$). The likelihood function is:

$$\tilde{f}_Z\big(z; O(s), P(s)\big) = \begin{cases} \pi(s) + (1-\pi(s)) \times e^{-\theta(s)}, & \text{if } z = 0 \\ (1-\pi(s)) \times \frac{\theta(s)^z e^{-\theta(s)}}{z!(1-e^{-\theta(s)})}, & \text{if } z > 0. \end{cases} . \tag{4.7}$$

Mixture Model for Spatial Semi-continuous Data

The zero-inflated Tobit (ZIT) model is an extension of mixture models for zero-inflated semi-continuous data. Similar to the ZIP model, both the occurrence $O(s)$

Table 4.1. Spatial two-part models broken down by class and observation type.

Class	Data Type	Occurrence $O(s)$	Prevalence $P(s)$
HURDLE	Discrete	Bernouilli	Zero-Truncated Poisson
		Bernouilli	Zero-Truncated Neg. Binomial
	Continuous	Bernouilli	Lognormal
		Bernouilli	Log skew-normal
Mixture	Discrete	Bernouilli	Poisson
		Bernouilli	Negative Binomial
	Continuous	Bernouilli	Tobit Model

and prevalence processes $P(s)$ generate zeros. The distribution for the prevalence process $P(s)$ is a Tobit model. The Tobit model generates censored observation $Z(s) \in \{0, \mathbb{R}^+\}$ as:

$$Z(s) = \begin{cases} Z^*(s), & \text{if } Z^*(s) > \gamma \\ 0, & \text{if } Z^*(s) \leq \gamma \end{cases}, \tag{4.8}$$

where γ is a threshold and $Z^*(s)$ is a latent random variable such that $Z^*(s) \sim \mathcal{N}(\mu(s), \tau^2(s))$. For the case where $\gamma = 0$, the Tobit model provides the following likelihood $f_Z\big(z; \theta(s)\big)$:

$$f_Z\big(z; \theta(s)\big) = \begin{cases} \Phi(\frac{\mu(s)}{\tau(s)}), & \text{if } z = 0 \\ \phi(\frac{z-\mu(s)}{\tau(s)}), & \text{if } z > 0. \end{cases}, \tag{4.9}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $\phi(\cdot)$ is the standard normal probability density function. Model parameters are $\theta(s) = \{\mu(s), \tau^2\}$. Consequently, the likelihood function for the zero-inflated Tobit (ZIT) model is:

$$\tilde{f}_Z\big(z; O(s), P(s)\big) = \begin{cases} \pi(s) + (1 - \pi(s)) \times \Phi(\frac{\mu(s)}{\tau(s)}), & \text{if } z = 0 \\ (1 - \pi(s)) \times \phi(\frac{z-\mu(s)}{\tau(s)}), & \text{if } z > 0. \end{cases}, \tag{4.10}$$

In Table 4.3.2, we outline the various two-part models for zero-inflated spatial data and provide details regarding the spatial processes.

4.3.3 Modeling and computational challenges

The occurrence and prevalence processes are modeled using spatial generalized linear mixed models (SGLMMs). Fitting SGLMMs to high-dimensional observations can be computationally costly due to large matrix operations and slow mixing random effects (Haran, 2011). SGLMMs typically require a costly evaluation of an $n-$dimensional multivariate normal likelihood function ($\mathcal{O}(n^3)$) at each iteration of the MCMC algorithm. Moreover, highly correlated spatial random effects can lead to poor mixing in MCMC algorithms (cf. Christensen et al., 2006; Haran et al., 2003).

Frequentist methods have been proposed for zero-inflated observations. These methods use two-stage procedures for Guass-Hermite quadrature (Min and Agresti, 2005), expectation-maximization (Lambert, 1992b; Roeder et al., 1999), or restricted maximum quasi-likelihoods (Kim et al., 2012). However, such approximations do not scale well with high-dimensional random effects. One study proposes a novel Markov Chain Maximum Likelihood (MCML) approach for modeling large zero-inflated spatial count data (Lyashevska et al., 2016), but this remains computationally demanding as it requires parallelized operations across multiple processors.

In the Bayesian framework, the literature focuses on improving the sophistication of zero-inflated spatial models. Studies have modeling spatio-temporal dependence (Fernandes et al., 2009; Neelon et al., 2016a; Arcuti et al., 2016), addressed overdispersion (Gschlößl and Czado, 2008; Lee et al., 2016), used skewed distributions (Dreassi et al., 2014; Liu et al., 2016), used t-distributions to model heavy tailed behavior (Neelon et al., 2015), and modeled prevalence with scale mixtures of normal distributions (Fruhwirth-Schnatter and Pyne, 2010). However, there is a dearth of research on addressing computational issues with large zero-inflated spatial datasets, namely the high-dimensional correlated spatial random effects and slow mixing of MCMC algorithms. For example, one study (Wang et al., 2014) models the presence and abundance of Atlantic cod in 1325 locations along the Gulf of Maine using predictive processes (Banerjee et al., 2008). Other studies provide methods to facilitate posterior sampling by representing the latent spatial processes as scale mixtures of normal distributions via dirichlet processes (Neelon et al., 2018) or through Pólya-Gamma mixtures (Neelon et al., 2019).

4.4 Computationally efficient zero-inflated spatial models

In this section, I we propose a computationally efficient approach for fitting two-part models for high-dimensional zero-inflated spatial data. Our method extends the projection-based intrinsic conditional autoregression (PICAR) framework to two=part models by imposing a non-parametric basis representation of the underlying spatial occurrence $O(s)$ and prevalence $P(s)$ processes. We present the hierarchical modeling framework and also provide practical guidelines for selecting tuning parameters. Finally, we examine the computational benefits afforded by the PICAR approach.

4.4.1 Projection Intrinsic Autoregression (PICAR)

We introduce a projection-based intrinsic conditional autoregression (PICAR) approach that is designed to efficiently fit hierarchical spatial models. In this framework, we represent spatial random effects $\mathbf{W} = (W(s_1), ..., W(s_n))$ as a linear combination of basis functions:

$$\mathbf{W} \approx \mathbf{\Phi}\delta \qquad , \qquad \delta \sim \mathcal{N}(0, \Sigma_\delta),$$

where $\mathbf{\Phi}$ is an $n \times p$ basis function matrix where each column denotes a basis function, $\delta \in \mathbb{R}^p$ are the re-parameterized spatial random effects (or basis coefficients), and Σ_δ is the $p \times p$ covariance matrix for the weights. Basis functions can be interpreted as a set of distinct spatial patterns that can be used to construct a spatial random field, along with their coefficients. Basis representation has been a popular approach to model spatial data (cf. Cressie and Johannesson, 2008; Banerjee et al., 2008; Hughes and Haran, 2013; Lindgren et al., 2011; Rue et al., 2009; Christensen et al., 2006; Haran et al., 2003; Griffith, 2003; Higdon, 1998; Nychka et al., 2015). Examples of basis functions include splines, wavelets, empirical orthogonal functions, combinations of sines and cosines, piece-wise linear functions, and many others. Basis representations tend to be computationally efficient as they help bypass large matrix operations, reduce the dimensions of the spatial random effects, and as in our case, decorrelate the spatial random effects

$\mathbf{W}$.

The Projection-based intrinsic conditional auto-regression (PICAR) approach consists of three components: (1) generate a triangular mesh on the spatial domain $\mathcal{D} \subset \mathbb{R}^2$; (2) construct a spatial field on the mesh vertices using non-parametric basis functions; (3) interpolate onto the observation locations using piece-wise linear basis functions. We provide additional details for each component.

Mesh Construction

Prior to fitting the model, we generate a mesh enveloping the observed spatial locations via Delaunay Triangulation (Hjelle and Dæhlen, 2006). Here, we divide the spatial domain D into a collection of non-intersecting irregular triangles. The triangles can share a common edge, corner (i.e. nodes or vertices), or both. The mesh generates a latent undirected graph $G = \{V, E\}$, where $V = \{1, 2, ..., m\}$ are the mesh vertices and E are the edges. Each edge E is represented as a pair (i, j) denoting the connection between i and j. The graph G is characterized by its weights matrix $\mathbf{N}$, an $m \times m$ matrix where $N_{ii} = 0$ and $N_{ij} = 1$ when mesh node i is connected to node j and $N_{ij} = 0$ otherwise. The triangular mesh is built using the **R-INLA** package (Lindgren et al., 2015). Guidelines for mesh construction are provided in Lindgren et al. (2015), and details pertaining to algorithms for Delaunay triangulation can be found in Hjelle and Dæhlen (2006).

Moran's Basis Functions

We generate a spatial random field on the set of mesh vertices V of graph G using the Moran's basis functions (Hughes and Haran, 2013; Griffith, 2003). Griffith (2003) propose an augmented spatial generalized linear mixed model using a subset of eigenvectors of the Moran's operator $(\mathbf{I} - \mathbf{11}'/m)\mathbf{W}(\mathbf{I} - \mathbf{11}'/m)$, where $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is a vector of 1's. Note that this operator is a component of the Moran's I statistic:

$$I(A) = \frac{m}{\mathbf{1}'\mathbf{W}\mathbf{1}} \frac{\mathbf{Z}'(\mathbf{I} - \mathbf{11}'/m)\mathbf{W}(\mathbf{I} - \mathbf{11}'/m)\mathbf{Z}}{\mathbf{Z}'(\mathbf{I} - \mathbf{11}'/m)\mathbf{Z}},$$

a diagnostic of spatial dependence (Moran, 1950) used for areal spatial data. Values of the Moran's I above $-\frac{1}{m-1}$ indicate positive spatial autocorrelation and values

below $-\frac{1}{m-1}$ indicate negative spatial autocorrelation (Griffith, 2003). Positive eigencomponents of the Moran's operator correspond to varying magnitudes and patterns of positive spatial dependence, or clustering. For the triangular mesh, the positive eigenvectors represent the patterns of spatial dependence among the mesh nodes, and their corresponding eigenvalues denote the magnitude of spatial dependence.

We construct the Moran's basis function matrix $\mathbf{M} \in \mathbb{R}^{m \times p}$, by selecting the first p eigenvectors of the Moran's operator where $p \ll m$. Rank selection for p proceeds via an automated heuristic (Lee and Haran, 2019) based on out-of-sample cross-validation. A spatial random field can be constructed through a linear combinations of the Moran's basis functions (contained in matrix $\mathbf{M}$) and their corresponding weights $\delta \in \mathbb{R}^p$.

Piece-wise Linear Basis Functions

To complete the PICAR approach, we introduce a set of piece-wise linear basis functions (Brenner and Scott, 2007) to interpolate points within the triangular mesh (i.e. the undirected graph $G = (V, E)$). We construct a spatial random field on the mesh nodes $\tilde{\mathbf{W}} = (W(v_1), ..., W(v_m))$ where $v_i \in V$ and then project, or interpolate, onto the observed locations $\mathbf{W} = (W(s_1), ..., W(s_n))$ where $s_i \in \mathcal{D}$. The latent spatial random field $\mathbf{W}$ can be represented as $\mathbf{W} = \mathbf{A}\tilde{\mathbf{W}}$, where $\mathbf{A}$ is an $n \times m$ projector matrix containing the piece-wise linear basis functions.

The rows of $\mathbf{A}$ correspond to an observation location $s_i \in \mathcal{D}$, and the columns correspond to a mesh node $v_i \in V$. The ith row of $\mathbf{A}$ contains the weights to linearly interpolate $W(s_i)$. To illustrate, when the observation location s_i is wholly contained within one of the mesh triangles, there will be three non-zero values in the ith row of the projector matrix $\mathbf{A}$, each corresponding to a mesh node $v_j \in V$. When the observation location lies on an edge between two mesh nodes, there will be two non-zero values in the corresponding row of $\mathbf{A}$. Finally, there will only be one non-zero value in the corresponding row when the observation location and mesh node share the same location. In practice, we use an $n \times m$ projector matrix $\mathbf{A}$ for fitting the hierarchical spatial model. For model validation and prediction, we generate an $n_{CV} \times m$ projector matrix $\mathbf{A}_{CV}$ that interpolates onto the n_{CV} validation locations.

PICAR Representation of Spatial Generalized Linear Mixed models (SGLMMs)

In the previous section, we introduced three major components of the PICAR approach: (1) the Moran's basis function matrix $\mathbf{M} \in \mathbb{R}^{m \times p}$; (2) the projector matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$; and (3) the corresponding weights $\delta \in \mathbb{R}^p$. Given a set of weights δ and the Moran's basis functions $\mathbf{M}$, we can build a spatial random field on the triangular mesh nodes $v \in V$ as $\tilde{\mathbf{W}} = \mathbf{M}\delta$, where $\tilde{\mathbf{W}} = (W(v_1), ..., W(v_m))$ for $v_i \in V$. Next, we linearly interpolate the latent spatial random field at the observation locations as $\mathbf{W} = \mathbf{A}\tilde{\mathbf{W}} = \mathbf{AM}\delta$, where $\mathbf{W} = (W(s_1), ..., W(s_n))$ for $s_i \in \mathcal{D}$, the spatial domain.

The PICAR approach can be embedded into the SGLMM modeling framework:

$$
\begin{aligned}
&\textbf{Data Model:} && Z(s)|\eta(s) \sim f(\eta(s)), \\
& && \eta(s) = g(E[Z(s)|\beta, \delta]) = X(s)\beta + [\mathbf{AM}\delta](s), \\
&\textbf{Process Model:} && \delta|\tau \sim \mathcal{N}(0, \tau^{-1}(\mathbf{M}'\mathbf{Q}\mathbf{M})^{-1}), \\
&\textbf{Parameter Model:} && \beta \sim N(\mu_\beta, \Sigma_\beta), \quad \tau \sim G(\alpha_\tau, \beta_\tau),
\end{aligned}
$$

where $\mathbf{A}$ is the projector matrix, $\mathbf{M}$ is the Moran's basis functions matrix, δ are the basis coefficients, $\mathbf{Q}$ is the prior precision matrix for the mesh vertices, τ is the precision parameter, and $\alpha_\tau, \beta_\tau, \mu_\beta$, and Σ_β are the hyperparameters.

By default, we set $\mathbf{Q}$ to be the precision matrix of an intrinsic conditional auto-regressive model (ICAR) fit on the mesh vertices V. Here, $\mathbf{Q} = (\text{diag}(\mathbf{N1}) - \mathbf{N})$, where $\mathbf{N}$ is the adjacency or weight matrix from Section 3.3.1 and $\mathbf{1}$ is m-dimensional vector of 1s. Since $\mathbf{Q}$ is not positive definite, this framework cannot be used within the likelihood function; however, it can be set as the prior distribution for the spatial random effects as part of the Bayesian hierarchical spatial model (Besag et al., 1991). Alternative options for $\mathbf{Q}$ and a comparative analysis can be found in Lee and Haran (2019).

4.4.2 PICAR Approach for Zero-inflated Spatial Data

The PICAR approach for fitting spatial generalized linear mixed models readily extends to the spatial two-part modeling framework (Equation 4.3.2). Here, we

replace the existing latent spatial random process ($\mathbf{W}_o$ and $\mathbf{W}_p$) with the PICAR representation $\mathbf{W}_o \approx \mathbf{A}_o\mathbf{M}_o\delta_o$ and $\mathbf{W}_p \approx \mathbf{A}_p\mathbf{M}_p\delta_p$. Here, $\mathbf{A}_o$ is a $n \times m$ projector matrix, $\mathbf{M}_o$ is a $m \times p_o$ matrix of Moran's basis functions, δ_o is a p_o-dimensional vector of basis coefficients for the occurrence process. Similarly, $\mathbf{A}_p$ is a $n_p \times m$ projector matrix, $\mathbf{M}_p$ is a $m \times p_p$ matrix of Moran's basis functions, and δ_p is a p_p-dimensional vector of basis coefficients for the prevalence process. The general hierarchical framework for the PICAR representation of two-part models is as follows:

Data Model:	$\mathbf{Z}\|O(s), P(s) \sim \tilde{f}_Z\big(z; O(s), P(s)\big)$
Process Model:	$O(s)\|\pi(s) \sim \text{Bern}(\pi(s))$
	$P(s)\|\theta(s) \sim f(\theta(s))$
Sub-process Model 1:	$\pi(s)\|\eta_o(s) = g_o^{-1}(\eta_o(s))$
(Occurrence)	$\eta_o(s)\|\beta_o, \delta_o(s), \epsilon_o(s) = X(s)\beta_o + \mathbf{A}_o\mathbf{M}_o\delta_o + \epsilon_o(s)$
	$\delta_o\|\tau_o \sim \mathcal{N}(\mathbf{0}, \tau_o^{-1}(\mathbf{M}_o'\mathbf{Q}_o\mathbf{M}_o)^{-1})$,
	$\epsilon_o(s)\|\tau^2_{\epsilon_o} \sim \mathcal{N}(0, \tau^2_{\epsilon_o})$
Sub-process Model 2:	$\theta(s)\|\eta_p(s) = g_p^{-1}(\eta_p(s))$
(Prevalence)	$\eta_p(s)\|\beta_p, W_p(s), \epsilon_p(s) = X(s)\beta_p + \mathbf{A}_p\mathbf{M}_p\delta_p + \epsilon_p(s)$
	$\delta_p\|\tau_p \sim \mathcal{N}(\mathbf{0}, \tau_p^{-1}(\mathbf{M}_p'\mathbf{Q}_p\mathbf{M}_p)^{-1})$,
	$\epsilon_p(s)\|\tau^2_{\epsilon_p} \sim \mathcal{N}(0, \tau^2_{\epsilon_p})$
Parameter Model:	Priors for $\beta_o, \tau_o, \tau^2_{\epsilon_o}, \beta_p, \tau_p,$ and $\tau^2_{\epsilon_p}$,

where $\tilde{f}_Z\big(z; O(s), P(s)\big)$ is the likelihood function of the zero-inflated data from Equation 4.4. The new components of the PICAR representation are projector matrices $\mathbf{A}_o$ and $\mathbf{A}_p$, the Moran's basis functions matrices $\mathbf{M}_o$ and $\mathbf{M}_p$, the basis coefficients δ_o and δ_p, and the precision parameters τ_o and τ_p. The $m \times m$ prior precision matrix for the mesh vertices $\mathbf{Q}$ is typically fixed prior to model fitting. Additional details are provided in the following section. Similar to the general framework from section 4.3.2, the data likelihood function $\tilde{f}_Z\big(z; O(s), P(s)\big)$ specifies the type of two-part models such as hurdle models (Equations 4.5 and 4.6) and mixture models (Equations 4.7 and 4.10).

4.4.3 Tuning Mechanisms

The occurrence $O(s)$ and prevalence $P(s)$ process models are based on spatial generalized linear mixed model (Section 4.3.1). The general SGLMM framework assumes that the true latent spatial random field $\mathbf{W} = \{W(s_1), W(s_2), ..., W(s_n)\}$ is a Gaussian process such that $\mathbf{W} \sim N(0, \sigma^2 R_\phi)$ with partial sill σ^2 and correlation matrix R_ϕ. On the other hand, the PICAR approach considers the latent spatial random field following a basis representation such that $\mathbf{W} \approx \mathbf{AM}\delta$, where $\delta \sim N(0, \tau^{-1}(\mathbf{M'QM})^{-1})$, $\mathbf{M}$ is the $m \times p$ Moran's basis function matrix, and $\mathbf{A}$ is the $n \times m$ projector matrix. An alternative formulation of the latent spatial random field is

$$\mathbf{W} \sim N(0, \tau^{-1}\mathbf{AM}(\mathbf{M'QM})^{-1}\mathbf{M'A'}).$$

The PICAR approach approximates the covariance matrix $\sigma^2 R_\phi$ such that $\sigma^2 R_\phi \approx \tau^{-1}\mathbf{AM}(\mathbf{M'QM})^{-1}\mathbf{M'A'}$. We proceed by tuning the rank of the Moran's operator rank($\mathbf{M}$) and the prior precision matrix $\mathbf{Q}$ of the mesh vertices.

The following automated heuristic selects the appropriate ranks p_o and p_p for the Moran's basis function matrices for both processes, $\mathbf{M}_o$ and $\mathbf{M}_p$. First, we generate two augmented datasets - $\mathbf{Z}_o^*$ and $\mathbf{Z}_p^*$- using the original zero-inflated spatial dataset $\mathbf{Z}$. The first dataset is generated as:

$$Z_o^*(s) = \begin{cases} 0, & \text{if } Z(s) = 0 \\ 1, & \text{if } Z(s) > 0. \end{cases}, \tag{4.11}$$

The second dataset $\mathbf{Z}_p^* \in \mathbb{R}^{n_p}$ is the collection of all observations such that $Z(s) > 0$ and n_p corresponds to the sample size of $\mathbf{Z}_p^*$. Next, we generate a set $\mathcal{P}$ consisting of h equally spaced points within the interval $[2, P]$ where P is the maximum rank and h is the interval resolution ($h = P - 1$ by default). Here, $P < m$ and both P and h are chosen by the user.

For the augmented dataset $\mathbf{Z}_o^*(s)$, we proceed in the following way. For each $p \in \mathcal{P}$, we construct an $n \times (k + p)$ matrix of augmented covariates $\tilde{X}_o = [X \quad \mathbf{A_o M_p}]$ where $X \in \mathbb{R}^{n \times k}$ is the original covariate matrix, $\mathbf{A} \in \mathbb{R}^{n \times m}$ is the projector matrix, and $\mathbf{M_p} \in \mathbb{R}^{m \times p}$ are the leading p eigenvectors of the Moran's operator. Next, we use maximum likelihood approaches to fit the appropriate generalized linear model (GLM) for binary responses, typically with a logit link function. Finally, we set

p_o to be the rank p that yields the lowest out-of-sample cross-validated root mean squared prediction error (CVRMSPE).

We implement a similar procedure for the second augmented dataset $\mathbf{Z}_p^* \in \mathbb{R}^{n_p}$. For each $p \in \mathcal{P}$, we construct an $n_p \times (k+p)$ matrix of augmented covariates $\tilde{X}_p = [X_p \quad \mathbf{A_p M_p}]$ where $X_p \in \mathbb{R}^{n_p \times k}$ is the matrix of covariates, $\mathbf{A}_{n_p} \in \mathbb{R}^{n_p \times m}$ is the projector matrix, and $\mathbf{M_p} \in \mathbb{R}^{m \times p}$ are the leading p eigenvectors of the Moran's operator. Note that the rows of X_p and $\mathbf{A}_p$ correspond to the n_p observations with positive values. Next, we use maximum likelihood approaches to fit the appropriate generalized linear model (GLM) for positive responses. For count data, the likelihood function is a zero-truncated Poisson distribution. For semi-continuous data in the HURDLE model framework, we employ a lognormal distribution as the likelihood function. For semi-continuous data in the mixture model framework, we simply fit the traditional linear model. Then, we set p_p to be the rank p that yields the lowest out-of-sample cross-validated root mean squared prediction error (CVMSPE).

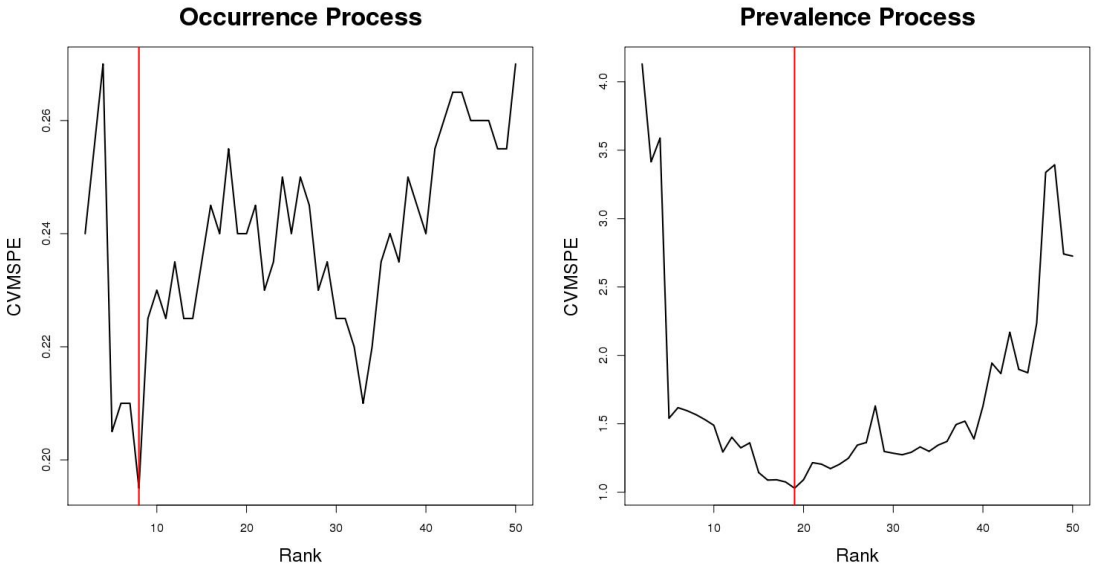

Figure 4.1. Cross-validated mean squared prediction error (CVMSPE) by ranks 1-50 using the automated heuristic for the occurrence (left) and prevalence (right) processes. The vertical red lines denote the chosen ranks ($p_o = 8$ and $p_o = 19$) with lowest CVMPSE.

Figure 4.1 demonstrates the automated heuristic for a spatial hurdle count model for a simulated zero-inflated spatial dataset from a HURDLE count model.

Note that ranks $p_o = 8$ and $p_p = 19$ results in the lowest CVMSPE for the occurrence and prevalence components.

Next, we provide some choices for $\mathbf{Q}$, the prior precision matrix for the mesh vertices $\tilde{\mathbf{W}}$. By default (Section 3.3.1), we set $\mathbf{Q}$ to be the precision matrix of an intrinsic conditional auto-regressive model (ICAR). Similarly, we could set $\mathbf{Q}$ as the precision matrix of a conditional auto-regressive model (CAR). Here, $\mathbf{Q} = (\mathbf{N1} - \rho\mathbf{N})$, where N is the adjacency matrix and $\rho \in (0, 1)$ is a predetermined correlation coefficient. It is possible to estimate ρ as a model parameter, but doing so requires an eigendecomposition of the Moran's operator ($\mathcal{O}(m^3)$) at each iteration of the MCMC algorithm, which can negate the computational gains of the PICAR approach. Another alternative is setting $\mathbf{Q} = I$, where the mesh nodes $\tilde{\mathbf{W}}$ and re-parameterized spatial random effects δ are uncorrelated.

4.4.4 Computational Advantages

The PICAR approach requires shorter computational times per iteration as well as fewer iterations for the Markov chain to converge. The computational speedup results from bypassing expensive matrix operations (e.g. Cholesky decomposition) and by decorrelating and reducing the dimensions of the spatial random effects. The computational cost is dominated by the matrix-vector multiplication $\mathbf{AM}\delta$, where $\mathbf{AM}$ is the $n \times p$ basis function matrix constructed prior to model fitting and δ are reparameterized spatial random effects (basis coefficients). The PICAR approach has a computational complexity of $\mathcal{O}(np)$ as opposed to $\mathcal{O}(n^3)$ for the full hierarchical spatial model.

We examine mixing in MCMC algorithms within the context of spatial generalized linear mixed models (SGLMMs). Here, the PICAR approach generates a faster mixing MCMC algorithm than fitting the full two-part model using the reparameterization method (gold standard) (Christensen et al., 2006). Note that this reparameterization approach is designed to improve mixing for in MCMC algorithms. This is corroborated by the larger effective sample size per second (ES/sec), the rate at which independent samples are generated by the MCMC algorithm. Larger values of ES/sec indicates faster mixing. In the simulated examples (Section 4.5), the PICAR approach returns a larger ES/sec than the gold

standard across all model parameters and spatial random effects. Additional details are provided in Section 4.5.

For PICAR, the two major computational bottlenecks are constructing the Moran's operator (Section 4.4.1) and computing its eigencomponents. The Moran's operator requires the matrix operation $(\mathbf{I} - \mathbf{1}\mathbf{1}'/m)\mathbf{W}(\mathbf{I} - \mathbf{1}\mathbf{1}'/m)$ and $2m^3 - m^2$ floating point operations (FLOPs), which may be computationally prohibitive for large datasets. We reduce computational costs by leveraging the embarrassingly parallel operations as well the sparsity of the weights matrix $\mathbf{N}$. We use the sparse matrix R package **Matrix** (Bates and Maechler, 2019) to reduce costs for the operation $\Sigma = (\mathbf{I} - \mathbf{1}\mathbf{1}'/m)\mathbf{N}$. Then, we partition the resulting matrix Σ into K mutually exclusive $\frac{1}{K} \times n$ sub-matrices Σ_k for $k = 1, ..., K$. By parallelizing across K processors, we can quickly construct the partial Moran's Operator $MO_k = \Sigma_k(I - 11'/m)$ for $k = 1, ..., K$. Finally, we generate the full Moran's Operator by combining the MO_k's as so $MO = \begin{bmatrix} MO_1 \\ \vdots \\ MO_K \end{bmatrix}$.

We can compute the k eigencomponents of the Moran's Operator using a partial eigendecomposition approach such as the Implicitly Restarted Arnoldi Method (Lehoucq et al., 1998) from **RSpectra** package (Qiu and Mei, 2019). Since the PICAR approach typically selects a rank$(\mathbf{M}) \ll n$, there is no need to perform a full eigendecomposition of the Moran's Operator $\mathbf{M}$.

4.5 Simulated Examples

We demonstrate our approach on simulated datasets generated from the following two-part models:

1. **Hurdle Model with Count Data:** The occurrence process $O(s)$ solely generates zero-values. The prevalence process $P(s)$ generates positive counts from a zero-truncated Poisson distribution.

2. **Hurdle Model with Semi-continuous Data:** The occurrence process $O(s)$ solely generates zero-values. The prevalence process $P(s)$ generates positive continuous values from a lognormal distribution.

3. **Mixture Model with Count Data:** The occurrence $O(s)$ and prevalence $P(s)$ processes both generate zero-values. The prevalence process $P(s)$ generates counts from a Poisson distribution.

4. **Mixture Model with Semi-continuous Data:** The occurrence $O(s)$ and prevalence $P(s)$ processes both generate zero-values. The prevalence process $P(s)$ generates counts from a Tobit model.

For each zero-inflated spatial model, we generate a dataset with locations on the unit domain $[0,1]^2$. Each dataset consists of 700 locations randomly chosen on the spatial domain. We use 500 observations to fit the hierarchical spatial model and reserve 200 observations for validation. We chose $n = 500$ so that we could compare against a gold standard method (below) for which higher dimensions can be computationally prohibitive. Each dataset includes a randomly generated matrix of covariates X. For the PICAR approach, rank selection proceeds via the automated heuristic presented in Section 4.4.3. We set the precision matrix of the occurrence process $\mathbf{Q}_o$ to be an $n_o \times n_o$ identity matrix and the precision matrix of the prevalence process $\mathbf{Q}_p$ to be an $n_p \times n_p$ identity matrix.

We compare our PICAR-based approach to fitting the full two-part model (gold standard). We fit the full two-part model using the reparameterization approach from Christensen et al. (2006), which is designed to improve mixing over default MCMC algorithms. Christensen et al. (2006) reparameterize the latent spatial process as $\mathbf{W}_o = \mathbf{L}_o\gamma_o$ and $\mathbf{W}_p = \mathbf{L}_p\gamma_o$ where $\mathbf{L}_o$ and $\mathbf{L}_p$ are the lower triangular matrices obtained through the Cholesky decomposition of the corresponding covariance matrices (i.e., $C_o(h;\sigma_o^2,\phi_o) = \mathbf{L}_o\mathbf{L}_o'$ and $C_p(h;\sigma_p^2,\phi_p) = \mathbf{L}_p\mathbf{L}_p'$). Next, γ_o and γ_p are reparameterized spatial random effects for the occurrence and prevalence processes, respectively. Additional details are provided in the appendix.

We examine the out-of-sample cross-validated root mean squared prediction error (CVRMSPE):

$$\text{CVRMSPE} = \sqrt{\frac{1}{n_{CV}}\sum_{i=1}^{n_{CV}}(Y_i^* - \hat{Y}_i^*)^2},$$

where $n_{CV} = 200$, Y_i^*'s denote the i-th value in the validation sample, and $\hat{Y}_i^*$'s

are the predicted values at the i-th location.

4.5.1 HURDLE Model for Spatial Count Data

In the first example, we simulate a zero-inflated spatial count dataset from a spatial hurdle model for count data. For both the occurrence $O(s)$ and prevalence $P(s)$ processes, $\beta_o = \beta_p = (1, 1)^T$, and the random effects ($\mathbf{W_o}$ and $\mathbf{W_p}$) are generated using the Matérn covariance function with $\nu = 2.5$, $\sigma^2 = 1$, and $\phi = 0.2$. The covariance function has the form (cf. Rasmussen and Williams, 2006)

$$C(h) = \sigma^2\Big(1 + \frac{\sqrt{5}|h|}{\phi} + \frac{5|h|^2}{3\phi^2}\Big)\exp\Big(-\frac{\sqrt{5}|h|}{\phi}\Big),$$

where $|h|$ is the Euclidean distance between locations. The data likelihood function follows Equation 4.5. The distribution for the prevalence process $f(\cdot|\theta(s))$ is a zero-Truncated Poisson distribution such that $f(\cdot|\theta(s)) = \frac{\theta(s)^z e^{-\theta(s)}}{z!(1-e^{-\theta(s)})}$. The occurrence process $O(s)$ is modeled as a spatial generalized linear mixed model (SGLMM) with a logit-link function $\text{logit}(\pi(s)) = \log\{\frac{\pi(s)}{1-\pi(s)}\}$. The prevalence process is also modeled as an SGLMM with a log-link function.

In the PICAR approach, we place a vague multivariate normal prior for the regression coefficients where $\beta_o \sim N(\mathbf{0}, 100I)$ and $\beta_p \sim N(\mathbf{0}, 100I)$. As in Hughes and Haran (2013), we chose a gamma prior for the precision parameters $\tau_o \sim G(0.5, 2000)$ and $\tau_p \sim G(0.5, 2000)$. For the binary and count data, we use Gibbs updates for τ_o and τ_p and random-walk Metropolis-Hastings updates for β_o and β_p. Finally, we update the basis coefficients δ_o and δ_p using an all-at-once Metropolis-Hastings update with a multivariate normal proposal centered at the parameters of the previous iterations as in Guan and Haran (2018). For the PICAR approach, we ran $100,000$ iterations of the MCMC algorithm. The MCMC algorithm is implemented using the programming language `nimble` (de Valpine et al., 2017).

The PICAR approach generates a triangular mesh with of $m = 528$ vertices. For the Moran's basis functions matrices, the automated heuristic selected ranks $p_o = 8$ and $p_p = 19$ for the occurrence and prevalence processes, respectively. Table 4.2 presents the parameter estimates, prediction results, and computational times for each simulated example. The PICAR approach provides similar point

and interval estimates to the gold standard method, as evidenced by the reported posterior means and the 95% credible intervals for regression parameters β_{1o}, β_{2o}, β_{1p}, and β_{1p}. Prediction results for both methods are similar where the PICAR approach yields a cross-validated root mean squared error (CVRMSPE) of 1.52 compared to a CVRMSPE of 1.68 for the gold standard. For this example, the PICAR approach outperforms the gold standard method in predictive ability, which is consistent with results from a past study that employs basis representations of spatial latent fields (Bradley et al., 2019). Moreover, the PICAR approach has a model fitting wall time of 1.2 minutes as opposed to 11.2 hours for the gold standard.

We also consider mixing in MCMC algorithms by examining the effective sample size per second (ES/sec), or the rate at which independent samples are generated by the MCMC algorithm. Here, larger values of ES/sec indicates faster mixing. The PICAR approach generates a faster mixing MCMC algorithm than the reparameterization method (Rep-SGLMM) (Christensen et al., 2006), an approach designed to improve mixing for SGLMMs. For model parameters β_{1o}, β_{2o}, β_{1p}, and β_{1p}, PICAR yields an ES/sec of 218.89, 214.15, 44.00, and 43.83, respectively. The gold standard returns an ES/sec 0.66, 0.63, 0.26, and 0.25, respectively. For the spatial random effects $\mathbf{W}_o(s)$ and $\mathbf{W}_p(s)$, the median ES/sec is 53.09 and 28.70 for the PICAR approach and 0.71 and 0.40 for the gold standard, an improvement by a factor of roughly 74.3 and 72.5. The computation times are based on a single 2.2 GHz Intel Xeon E5-2650v4 processor. All the code was run on the Pennsylvania State University Institute for CyberScience-Advanced CyberInfrastructure (ICS-ACI) high-performance computing infrastructure.

Example	Method	β_{1o}	β_{2o}	RMPSE	Time
HURDLE Count	Gold	1 (0.61,1.38)	1.19 (0.8,1.57)	1.68	11.2 hr
HURDLE Count	PICAR	1.02 (0.64,1.41)	1.27 (0.89,1.67)	1.52	1.2 min
HURDLE Semi	Gold	1.19 (0.77,1.64)	1.48 (1.06,1.92)	2.03	11.0 hr
HURDLE Semi	PICAR	1.1 (0.7,1.51)	1.49 (1.09,1.9)	2.01	1.2 min
Mix Count	Gold	1.25 (0.69,1.84)	0.85 (0.26,1.42)	1.66	14.5 hr
Mix Count	PICAR	1.48 (0.81,2.18)	0.95 (0.29,1.58)	1.77	1.6 min
Mix Semi	Gold	2.46 (1.38,3.55)	2.27 (1.17,3.38)	0.49	14.9 hr
Mix Semi	PICAR	1.12 (-0.94,3.1)	0.41 (-1.84,2.51)	0.52	3.5 min

Table 4.2. Inference, prediction, and computational results for simulated examples.

4.5.2 HURDLE Model for Spatial Semi-continuous Data

Next, we simulate a zero-inflated spatial semi-continuous dataset using a spatial hurdle model for semi-continuous data. The data is generated using the same specifications as in the HURDLE count case. However, the distribution for the prevalence process $f(z|\theta(s))$ is a lognormal distribution such that $f(z|\theta(s)) = \frac{1}{z\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\ln z - \mu(s))^2}{2\tau^2}\right)$ with parameters $\theta(s) = \{\mu(s), \tau^2\}$. We define $\mu(s) = X(s)\beta_p + W(s)$ and set the true value for $\tau^2 = 0.1$. The data likelihood function follows Equation 4.6. The occurrence process is modeled similar to the HURDLE count case, while the prevalence process adheres to a SGLMM with a lognormal likelihood function and a link function $g(\cdot) = exp(X(s)\beta_p + W(s) + \tau^2/2)$. We follow the specifications for model fitting as outlined in the HURDLE count case; however, we also estimate the variance of the lognormal distribution τ^2 and set prior distribution $\tau^2 \sim \text{IG}(2, 2)$.

In the PICAR approach, the triangular mesh consists of $m = 531$ vertices. For the Moran's basis functions matrices, the automated heuristic selected ranks $p_o = 5$ and $p_p = 39$ for the occurrence and prevalence processes, respectively. The PICAR approach provides similar point and interval estimates to the gold standard method as shown in Table 4.2. Prediction results for both methods are similar where the PICAR approach yields a cross-validated root mean squared error (CVRMSPE) of 2.01 compared to a CVRMSPE of 2.03 for the gold standard.

Moreover, the PICAR approach has a model fitting wall time of 1.2 minutes as opposed to 11.0 hours for the gold standard. Furthermore, the PICAR approach exhibits faster mixing than the gold standard as shown by the larger effective samples per second (ES/sec). For model parameters β_{1o}, β_{2o}, β_{1p}, β_{1p}, and τ^2 PICAR yields an ES/sec of 213.54, 223.32, 516.04, 76.83, and 26.18 and the gold standard returns an ES/sec of 0.44, 0.57, 1.52, 1.38, and 0.94, respectively. For the spatial random effects $\mathbf{W}_o(s)$ and $\mathbf{W}_p(s)$, the median ES/sec is 70.7 and 21.5 for the PICAR approach and 0.64 and 1.49 for the gold standard, an improvement by a factor of roughly 110.2 and 14.4.

4.5.3 Mixture Model for Spatial Count Data

The next dataset comes from a popular spatial mixture model for count data, the zero-inflated Poisson (ZIP) model. The model parameters follow those from the previous cases. For the ZIP model, the distribution for the prevalence process $f(z|\theta(s))$ is a Poisson distribution such that $f(z|\theta(s)) = \frac{\theta(s)^z e^{-\theta(s)}}{z!}$ with rate parameter $\theta(s)$. The data likelihood function follows Equation 4.7. The occurrence process is modeled as in the previous cases, and the prevalence process is modeled as an SGLMM with a log link function $g(\theta) = log(\theta(s))$. Model fitting for both the PICAR and gold standard cases area similar to the previous two cases.

The PICAR approach generates a triangular mesh with of $m = 526$ vertices. For the Moran's basis functions matrices, the automated heuristic selected ranks $p_o = 13$ and $p_p = 6$ for the occurrence and prevalence processes, respectively. The PICAR approach provides similar point and interval estimates to the gold standard method as shown in Table 4.2. Prediction results for both methods are similar where the PICAR approach yields a cross-validated root mean squared error (CVRMSPE) of 1.77 compared to a CVRMSPE of 1.66 for the gold standard. However, the PICAR approach has a model fitting wall time of 1.6 minutes as opposed to 14.5 hours for the gold standard. Furthermore, the PICAR approach exhibits faster mixing than the gold standard as shown by the larger effective samples per second (ES/sec). For model parameters β_{1o}, β_{2o}, β_{1p}, and β_{1p}, PICAR yields an ES/sec of 51.26, 60.98, 25.38, and 30.93, respectively. The gold standard returns an ES/sec 0.14, 0.11, 0.05, and 0.05, respectively. For the spatial random effects $\mathbf{W}_o(s)$ and $\mathbf{W}_p(s)$, the median ES/sec is 21.9 and 22.5 for the PICAR approach and 0.09 and 0.10 for the gold standard, an improvement by a factor of roughly 234.1 and 225.8.

4.5.4 Mixture Model for Spatial Semi-continuous Data

Finally, we examine a zero-inflated spatial semi-continuous dataset generated using a spatial mixture model for semi-continuous data. Here, the distribution for the prevalence process $f(z|\theta(s))$ is the Tobit model from Equation 4.9 with parameters $\theta(s) = \{\mu(s), \tau^2\}$. We define $\mu(s) = X(s)\beta_p + W(s)$ and set the true value for $\tau^2 = 0.1$. The data likelihood function follows Equation 4.10. We model the

occurrence process $O(s)$ similar to the previous cases. The prevalence process $P(s)$ is modeled according to an SGLMM driven by a Tobit model. We follow the specifications for model fitting as outlined in the HURDLE count case; however, we also estimate the variance of the Tobit model τ^2 and set prior distribution $\tau^2 \sim \text{IG}(2, 2)$.

For the PICAR approach, the triangular mesh includes $m = 531$ vertices. For the Moran's basis functions matrices, the automated heuristic selected ranks $p_o = 10$ and $p_p = 36$ for the occurrence and prevalence processes, respectively. As shown in Table 4.2, Prediction results for both methods are similar where the PICAR approach yields a cross-validated root mean squared error (CVRMSPE) of 0.52 compared to a CVRMSPE of 0.49 for the gold standard. The PICAR approach has a model fitting wall time of 3.5 minutes as opposed to 14.9 hours for the gold standard. Similar to the previous cases, the PICAR approach exhibits faster mixing than the gold standard as shown by the larger effective samples per second (ES/sec). For model parameters β_{1o}, β_{2o}, β_{1p}, and β_{1p} PICAR yields an ES/sec of 21.99, 14.96, 17.05, 14.88, and 13.63 and the gold standard returns an ES/sec 0.08, 0.07, 0.08, 0.05, and 0.06, respectively. For the spatial random effects $\mathbf{W}_o(s)$ and $\mathbf{W}_p(s)$, the median ES/sec is 13.1 and 12.5 for the PICAR approach and 0.09 and 0.07 for the gold standard, an improvement by a factor of roughly 144.7 and 177.7.

4.6 Application: Abundance of Bivalve Species in the Dutch Wadden Sea

We showcase the scalability of our method by modeling a high-dimensional ecological dataset. The region of interest is the Dutch Wadden sea, a protected ecological habitat made up of sand barriers, salt marshes, mudflats, and gullies (Compton et al., 2013; Lyashevska et al., 2016). The Dutch Wadden sea is a famous stopover site for shorebirds (Lyashevska et al., 2016), particularly due to the presence of the Baltic tellin (*Macoma balthica*), a species of benthic invertebrates. Here, we examine spatial abundance data of the *Macoma balthica* species from Lyashevska et al. (2016) originally obtained from the synoptic intertidal benthic surveys (SIBES)

monitoring program (Compton et al., 2013; Bijleveld et al., 2012). The observations consist of counts of the Baltic tellin (*Macoma balthica*) species sampled at $n = 4,029$ locations. Here, 65.9% of the locations have zero-counts. The occurrence (presence vs. absence) and prevalence (values of positive counts) maps are provided in Figure 4.6.

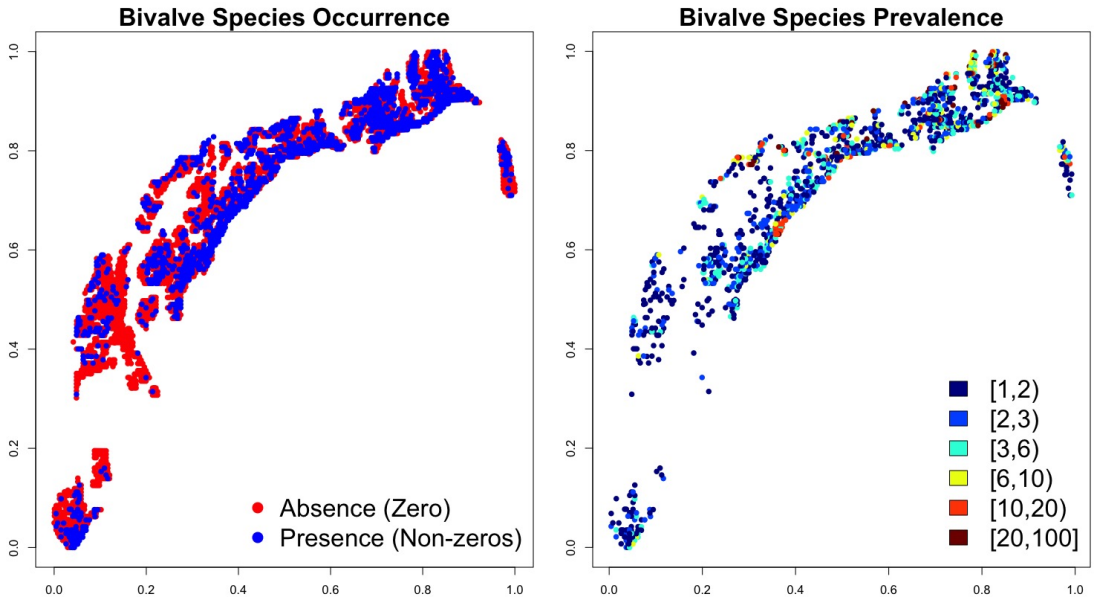

Figure 4.2. Maps of occurrence (left) and prevalence (right) of the Baltic tellin (Macoma balthica) species. For the occurrence map, the blue points denote the presence and the red points denote absence of the bivalve species. The prevalence map displays counts at the locations with positive counts.

We randomly select $3,220$ observations to fit our model and hold out $2,543$ observations for validation. Covariates include environmental variables that affect the abundance of the *Macoma balthica* species such as: (1) median grain size of the sediments; (2) silt content of the sediments; and (3) altitude. Using the PICAR approach, we fit the HURDLE count model and the zero-inflated Poisson model (mixture). We construct a triangular mesh with $m = 5,560$ mesh vertices, and the automated heuristic (Section 4.4.3) chose ranks $p_o = 70$ for the occurrence process $O(s)$ and $p_p = 10$ for the prevalence process $P(s)$.

We employ similar model specifications and prior distributions as in the simulated examples (Section 4.5), and we run the MCMC algorithms for $100,000$ iterations. Wall times for the PICAR approach were 10.3 minutes to fit the HURDLE

count model and 17.0 minutes to fit the zero-inflated Poisson (mixture) model. Comparisons to the gold standard approach (Christensen et al., 2006) are computationally prohibitive due to the long wall times associated with the MCMC algorithms. For instance, the gold standard approach for a HURDLE count model would require 5.6 months to run the same number of iterations of the MCMC algorithm as our method. Similarly, the zero-inflated Poisson (mixture) model version would need 12.35 months to do the same. We present the inference and prediction results in Table 4.6. Despite having a longer model fitting wall time, the zero-inflated Poisson (mixture) model exhibits better predictive ability over the HURDLE count model as evidenced by the lower root mean squared prediction error.

Process	Covariate	Estimate (95% CI) HURDLE Model	Estimate (95% CI) ZIP Model
Occurrence	Median grain size	-0.02 (-0.22 , 0.19)	-0.14 (-0.39 , 0.11)
Occurrence	Silt content	0.16 (-0.02 , 0.34)	-0.04 (-0.26 , 0.16)
Occurrence	Altitude	0.49 (0.4 , 0.59)	0.45 (0.34 , 0.57)
Prevalence	Median grain size	0.28 (0.19 , 0.36)	0.27 (0.18 , 0.35)
Prevalence	Silt content	0.68 (0.61 , 0.75)	0.68 (0.61 , 0.75)
Prevalence	Altitude	0.27 (0.24 , 0.31)	0.27 (0.23 , 0.3)
	RMSPE	4.6	4.4
	Wall Time	10.3 min	17.0 min

Table 4.3. Real Data Example: Inference and prediction results for the PICAR representation of the HURDLE count and zero-inflated Poisson (mixture) models. We provide the parameter estimates and 95% credible intervals for the regression coefficients corresponding to the three covariates (mean grain size, silt content, and altitude) and two processes (occurrence and prevalence). This includes prediction results (root mean squared prediction error) and model fitting wall times.

4.7 Discussion

In this study, we present a computationally efficient approach to model high-dimensional zero-inflated spatial observations. We modify the spatial two-part modeling framework by using a PICAR representation of latent spatial processes (occurrence $O(s)$ and prevalence $P(s)$). Our approach dramatically reduces model-fitting wall times while preserving inferential and predictive ability. We show that

our approach is computationally efficient, scales up to higher dimensions, automated, and extends to a variety of spatial two-part models for zero-inflated data. Moreover, our method can be readily implemented in a programming language for Markov chain Monte Carlo algorithms such as `nimble`. As shown in the simulated examples, our approach yields comparable results to the gold standard in both inference and prediction, but incurs a fraction of the computational costs. By drastically reducing computational cost, we are able to model high-dimensional spatial datasets such as the *Macome balthica* species abundance data within the order of minutes, as opposed to months for existing models.

Our method is subject to the following caveats. In this study, we focus on four commonly used two-part models. A natural extension would consider complex two-part models such as HURDLE models with skewed distributions (Dreassi et al., 2014; Liu et al., 2016), t-distributions to model heavy tailed behavior (Neelon et al., 2015), or scale mixtures of normal distributions (Fruhwirth-Schnatter and Pyne, 2010). Another avenue for future research would be to extend the PICAR approach to cases where there is correlation between the occurrence and prevalence processes (Recta et al., 2012). Finally, our approach does not provide an procedure for choosing between HURDLE and mixture models prior to model-fitting. Developing a pre-model-fitting test or automated heuristic would be a promising area of future research.

Chapter 5

Discussion and Future Work

My dissertation focuses on developing computationally efficient statistical methods for calibrating complex computer models and fitting high-dimensional hierarchical spatial models. The proposed methods address two key scientific problems: (1) incorporating information from various data sets to tune, or "calibrate" complex computer models to enable a better understanding of the past, present, and the future of the climate; and (2) developing fast algorithms to model complex spatial datasets from the environmental sciences. I apply these novel methods to real world examples in the environmental sciences such as ice sheet model calibration, ecological data analysis, and public health studies.

5.1 Summary and Contributions

In Chapter 2, I present a fast particle-based approach for calibrating complex computer models, namely a three-dimensional Antarctic ice sheet model. Complex ice sheet computer models play a prominent role in climate science, particularly in projecting future climate. These models require parameters that are calibrated based on observations and prior knowledge. As the number of parameters to be calibrated and as model complexity increases, current calibration methods either become computationally prohibitive or largely underestimate parametric uncertainty. Our approach employs a sequential Monte Carlo method that takes advantage of the massive parallelization afforded by modern high performance computing systems. The drastic reduction in computational times enables us to provide new insights

into important scientific questions, for example, the impact of Pliocene era data and prior parameter information on sea level projections. These studies would be computationally prohibitive with other computational approaches for calibration such as Markov chain Monte Carlo or emulation-based methods. I also find considerable differences in the distributions of sea level projections when we account for a larger number of uncertain parameters. This work provides a promising step forward towards improving the uncertainty quantification of complex, computationally intensive, and decision-relevant models.

In Chapter 3, I introduce a projection intrinsic conditional autoregression (PICAR) approach for modeling high-dimensional hierarchical spatial models. Hierarchical spatial models are commonly used across many fields such as ecology, glaciology, public health studies and criminology. However, these models pose considerable computational challenges due to the large number of highly correlated spatial random effects, which results in slow mixing Markov Chain Monte Carlo (MCMC) algorithms and expensive large matrix operations. I propose a computationally efficient approach for fitting high-dimensional hierarchical spatial models by de-correlating and reducing the dimensions of the spatial random effects. The major advantage of our method is that it is easily accessible for non-experts to specify general hierarchical spatial models of their choice and also provides an efficient estimation approach even for large datasets. More specifically, our approach: (i) automatically implements dimension reduction and decorrelation of the random effects; (ii) can be easily integrated into a hierarchical modeling scenario, as shown by our implementation in the language `Stan` and `NIMBLE`; and (iii) our method scales well to higher dimensional hierarchical spatial models. I demonstrate the PICAR approach on several simulated examples as well as two high-dimensional real-world spatial datasets.

In Chapter 4, I present a fast computational approach for modeling high-dimensional zero-inflated spatial data. Zero-inflated spatial data are common in many fields such as the climate sciences, ecology, public health, and epidemiology. Two-part models are commonly used to model zero-inflated spatial observations. However, fitting two-part models to large datasets can be computationally burdensome due to the high number of correlated spatial random effects. I modify the spatial two-part modeling framework by using a PICAR representation of latent

spatial processes (occurrence $O(s)$ and prevalence $P(s)$). This approach is computationally efficient, extends to a variety of spatial two-part models, and is readily implemented in a programming language for Markov chain Monte Carlo algorithms such as `nimble`. I demonstrate this approach on simulated and real-world datasets to showcase its computational efficiency and predictive ability.

In summary, the computational methods presented in this dissertation allows researchers to perform tasks that were previously infeasible such as calibrating a class of computer models and model certain types of high-dimensional spatial data. The particle-based calibration approach enables computer experiments that were computationally prohibitive. The PICAR approach models high-dimensional spatial data faster than gold standard approaches, but preserves inferential and predictive ability. This approach applies to a wide array of spatial models (e.g. spatially varying coefficients, ordinal spatial data, and zero-inflated models). Moreover, this approach can be readily incorporated into inference for a large number of user-specified hierarchical spatial models.

5.2 Caveats and Potential Improvements

There are several important caveats related to the computational methods developed in this dissertation. The fast particle-based approach introduced in Chapter 2 is designed for a specific class of computer models with moderate single mode run times (5 seconds - 15 minutes) and moderate number of model parameters (5 - 20). The particle-based approach may not be appropriate for computer models with long model run times ($>$ 15 minutes). Since this approach relies on massive parallelization (2000+ cores), even small increases in model run times can lead to a dramatic rise in computational costs. For these computer models, one may consider parallel MCMC algorithms where the fast mixing yields shorter mutation stages (fewer sequential model runs). Examples include multiple-try Metropolis (Martino, 2018; Liu et al., 2000) or ensemble Markov chain Monte Carlo (Neal, 2011). High-dimensional input spaces ($>$ 20 parameters) may also be problematic because the particle-based approach would require: (1) a large number of particles to sensibly approximate the target distribution; (2) longer mutation stages to move the particles into the high-probability regions; and (3) a massive allocation

computational resources. Past theoretical work (Crisan and Doucet, 2000) state that using more particles yields better approximations of the target distributions. In Lee et al. (2020), we chose the number of particles based on the available computational resources. Subsequent research would focus on selecting the optimal number of particles for particle-based calibration.

A number of caveats apply to our scientific findings. The PSU3D-ICE model runs at a coarser resolution than previous studies (DeConto and Pollard, 2016; Chang et al., 2016a,b; Pollard et al., 2016), which is a compromise between physical fidelity and run-time feasibility. At coarser resolutions, complex ice processes may not properly coalesce due to the spatial constraints. Replicating this calibration study at sharper spatial resolutions (40 km to 10 km) would be a fruitful extension of this study. Here, we would need to develop novel methods to further reduce sequential model evaluations in the mutation stage. Promising avenues for future work would include incorporating parallel MCMC approaches such as Multiple-Try Metropolis (Liu et al., 2000) or "emcee" samplers (Goodman and Weare, 2010), which enables faster mixing Markov chains and shorter mutation stages. Finally, the likelihood functions for the paleoclimate records may heavily influence calibration results. We have shown how the choice of expert priors influence sea level rise rojections, but the influence of likelihood functions remains an open area of research.

In Chapter 3, I present a computationally efficient approach for fitting high-dimensional hierarchical spatial models. The PICAR approach is subject to computational challenges for fitting hierarchical spatial models with a massive number of spatial random effects $100k+$. The dominating computational costs comes from an expensive eigendecomposition of the Moran's operator and repeated matrix-vector multiplications within the MCMC algorithm. Approximate eigendecomposition approaches such such as Nyström method (Williams and Seeger, 2001) or random projections (Banerjee et al., 2013; Guan and Haran, 2018) can alleviate some of these computational demands. Moreover, the repeated matrix-vector multiplications can be parallelized across multiple cores; thereby permitting the PICAR approach to scale up ultra-high-dimensional hierarchical spatial models. Another interesting possibility is incorporating the Moran's basis function into the latent conjugate multivariate (LCM) model framework (Bradley et al., 2019),

which could help bypass expensive Metropolis-Hastings updates. These caveats also extend to the PICAR representation of two-part models in Chapter 4

In Chapter 4, I focus on a subset of two-part models for zero-inflated spatial data. The proposed method does not explicitly consider the many other sophisticated two-part models such as HURDLE models with skewed distributions (Dreassi et al., 2014; Liu et al., 2016), t-distributions to model heavy tailed behavior (Neelon et al., 2015), or scale mixtures of normal distributions (Fruhwirth-Schnatter and Pyne, 2010). Another avenue for future research would be to extend the PICAR approach to cases where there is correlation between the occurrence and prevalence processes (Recta et al., 2012). Finally, our approach does not provide an procedure for choosing between HURDLE and mixture models prior to model-fitting. Developing a test or automated heuristic that proceeds before model-fitting would be a promising area of research.

5.3 Avenues for Future Research

Directions for future research stem from many of the caveats listed above. Here, I present five promising avenues for future research.

5.3.1 Computer Model Calibration for High-dimensional Spatial Binary Outputs

Recent calibration studies (Chang et al., 2016a,b) provide sharp sea level projections by assimilating high-dimensional non-Gaussian spatial observations of the Antarctic ice sheet (Fretwell et al., 2012). The higher resolution (40 km) PSU3D-ICE model generates a spatial field $n = 19600$ of binary (ice vs. no ice) outputs. The particle-based approach from Chapter 2 is not computationally feasible for this setting.

I propose using low-dimensional mismatch statistics obtained from the high-dimensional spatial model outputs and observations. My objective is to distill the discrepancy between the model output and observations into a single mismatch statistic. Summary statistics have been used in many Approximate Bayesian Computation (ABC) algorithms (Joyce and Marjoram, 2008; Wegmann et al., 2009;

Drovandi et al., 2011) and have also been used to build Gaussian process emulators for gravity models in epidemiology (Jandarov et al., 2014).

One example is a summary statistic that captures the spatial mismatch of the perimeter locations. Based on exploratory analysis, the PSU3D-ICE model consistently models ice presence in locations near the center of the ice sheet, independent of parameter settings. However, there is much more variability around the edge of the ice sheet. Consider mismatch locations, or knots, $s_{ice} \subset \mathcal{S}_{ice}$ and $s_{none} \subset \mathcal{S}_{none}$ where $\mathcal{S}_{ice}$ is the collection of m spatial points that lie 100 km inside the Antarctic ice sheet perimeter and $\mathcal{S}_{none}$ is the collection of m spatial points that lie 100 km outside the perimeter. Let $\mathcal{S} = \mathcal{S}_{ice} \cup \mathcal{S}_{none}$, which contains $2m$ spatial points.

The observations $Z(s)$ are defined as:

$$Z(s) = \begin{cases} 1 & \text{if } s \in \mathcal{S}_{ice} \\ 0 & \text{if } s \in \mathcal{S}_{none} \end{cases},$$

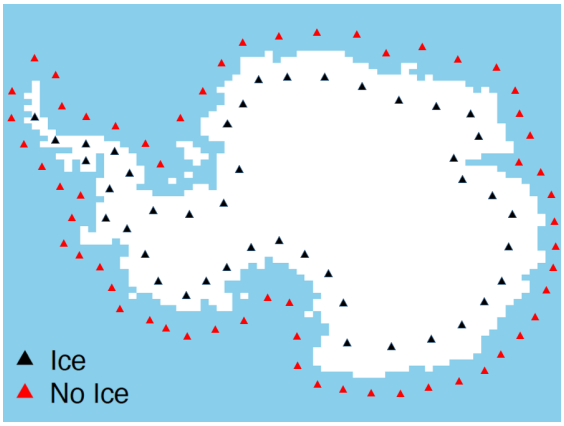

Figure 5.1. Map of the Antarctic ice sheet with mismatch locations, or knots. Blue triangles denote the locations with ice and lie 100 km inside the Antarctic ice sheet perimeter. Red triangles denote the locations without ice and lie 100 km outside the perimeter.

Consider the mismatch statistic $\tilde{Z}(s) = \sum_{\mathcal{S}} \mathbb{I}\{Z(s) = Y(s|\theta)\}$, where $Z(s)$ is the observation at location s and $Y(s|\theta)$ is the ice sheet model output at location s generated using parameter setting θ. This mismatch statistic represents the amount of spatial mismatch between the model output and observations around the edge of the ice sheet.

In a very simple case, $\tilde{Z}(s)$ has the following hierarchical structure:

$$\begin{aligned} \tilde{Z}(s) &\sim \text{Bin}(2m, p) \\ p &\sim \text{Beta}(\alpha, \beta), \end{aligned} \tag{5.1}$$

where α and β are hyperparameters for the prior Beta distribution. Note that by setting the prior distribution for p to be $\text{Beta}(10, 1)$, the model will heavily favor perfect matches between the model output and observations. Setting the prior to be $Beta(1, 1)$, a uniform prior on the unit scale, allows more flexibility for the spatial match. In this example, I fix the total number of trials to be $2m$.

5.3.2 Parallel MCMC approaches for Model Calibration

High-resolution (40 km or 20 km) ice sheet models incur heavy computational costs during the mutation stage of the particle-based algorithm due to their long model run times. In the mutation stage, adding a second layer of paralellization would reduce the number of sequential model runs. The idea is to propose multiple candidates for each particle, as opposed to one candidate in the particle-based approach. Then, fewer mutation updates should be necessary per iteration of the algorithm. The new approach would build upon parallelized Markov chain Monte Carlo methods, such as multiple-try Metropolis (Martino, 2018; Liu et al., 2000), locally weighted Markov chain Monte Carlo (Bernton et al., 2015), and ensemble Markov chain Monte Carlo (Neal, 2011). Note that adding a second layer of parallelization requires complex code development and coordination within the high performance computing system. Another viable option is to reduce the number of mutation updates is to resample particles from the entire mutation sampling paths (Nguyen, 2014; Del Moral et al., 2006; Gramacy et al., 2010); thereby providing a larger pool for resampling.

5.3.3 Extensions to Non-stationary and Spatio-temporal Models

Many applications address spatial data with non-stationary latent spatial processes or spatial processes that also exhibit temporal dependence. Developing computa-

tionally efficient approaches for these types of models has been an active area of research. Extending the PICAR approach to non-stationary hierarchical models or spatio-temporal hierarchical models would be challenging problem and potentially a useful contribution to the literature.

Non-stationary Hierarchical Spatial Models

In this dissertation, I focus on spatial models that assume the latent spatial random processes are weakly stationary, or the correlation between locations is a function of distance. On the other hand, non-stationary spatial processes are characterized by spatial dependence structures that tend to vary as a function of distances. Past studies have examined non-stationary spatial models using process convolutions (Higdon et al., 1999), non-stationary covariance functions (Paciorek and Schervish, 2006), averaging (smoothing) locally stationary processes (Fuentes, 2001; Nychka et al., 2018), and mapping locations to a deformed space where stationarity exists (Schmidt and O'Hagan, 2003). As sample size increases, existing approaches face considerable computational challenges due to the high-dimensional and heavily correlated spatial random effects.

I plan on extending the PICAR representation of latent spatial processes to non-stationary hierarchical spatial models. This includes spatial generalized linear mixed models as well as other complex spatial models with multiple latent processes. The objective is to develop a computationally efficient basis representation approach that scales up to high-dimensional non-stationary hierarchical spatial models.

One possible method would build upon the approach presented in Nychka et al. (2018). First, I plan on dividing the spatial domain into smaller overlapping regions. For each region, I would assign a localized set of Moran's basis functions $\mathbf{M}_i$ for regions $i = 1, ..., N$. Here, the basis functions will have 0's for locations outside the corresponding regions. In addition, I will specify a global set of Moran's basis functions $\mathbf{M}_g$. The combined set of Moran's basis functions $M_{\text{tot}} = [\mathbf{M}_1, ..., \mathbf{M}_N, \mathbf{M}_g]$ will be incorporated into the general hierarchical spatial modeling framework. While this proposed approach is shares similarities to the method from Nychka et al. (2018), my approach bypasses knot selection for the radial basis functions and uses the PICAR representation of the spatial random

field.

Spatio-temporal Models

Modern data collection initiatives have led to larger and more complex spatio-temporal datasets. Modeling high-dimensional spatio-temporal data faces considerable modeling and computational challenges. As the spatial or temporal resolution increases, storage and matrix operations of the resulting covariance matrices become computationally costly. These computational challenges affect models with non-separable spatio-temporal covariance structures, where the spatio-temporal covariance functions cannot be represented as a Kronecker product of the separate spatial and temporal covariance functions.

Past studies employ spatio-temporal basis functions (Wikle et al., 2019; Bryn-jarsdóttir and Berliner, 2014, cf.) which help circumvent the prohibitively large matrix operations. One promising avenue would use tensor products of spatial and temporal basis functions to represent the underlying spatio-temporal processes. However, constructing spatio-temporal basis functions is non-trivial as there are numerous spatial basis functions (see point below) and temporal basis functions. For future studies, I aim to construct computationally efficient spatio-temporal basis functions with a focus on selecting the appropriate spatial and temporal components.

5.3.4 A Comparative Study: Basis Functions for Spatial Non-Gaussian Data

Though basis representation methods have been widely used, the choice of basis functions are often left to the user. Examples of spatial basis functions include (1) bi-square (radial) basis functions (Cressie and Johannesson, 2008) with varying resolutions (Katzfuss, 2017; Nychka et al., 2015); (2) empirical orthogonal functions (Cressie and Wikle, 2015), or spatial representations of principal component analysis; (3) predictive process basis functions (Banerjee et al., 2008); (4) Moran's basis functions for spatial eigenfiltering (Hughes and Haran, 2013; Griffith, 2003); (5) piecewise Linear functions on a triangulation of the spatial domain (Lindgren et al., 2011); (6) square roots of the correlation matrix of the spatial random

effects via cholesky factorization Christensen et al. (2006) or approximate eigen-decompositions (Banerjee et al., 2013; Guan and Haran, 2018); (7) W-wavelets (Shi and Cressie, 2007) and multiresolution wavelet basis functions (Nychka et al., 2002); (8) Fourier basis functions composed of sine and cosine curves (Royle and Wikle, 2005); and (9) Gaussian kernel basis functions (Higdon, 1998). Heaton et al. (2019), Bradley et al. (2016), and Sun et al. (2012) compare a subset of these basis functions (along with other fast modeling approaches). However, these studies primarily focus on linear spatial models with Gaussian observations.

I propose a comparative study among spatial basis functions within the context of modeling non-Gaussian spatial observations. The goal is to provide clear guidelines for choosing the appropriate spatial basis functions depending on the observed data. A possible simulation study would examine non-Gaussian spatial datasets generated under the following scenarios:

- **Smoothness of the latent random field:** Vary the smoothness of the underlying spatial random field. For the Matérn class, the smoothness parameter would vary such that $\nu = 0.5$ (exponential), $\nu = 2.5$, and $\nu = \infty$ (squared exponential).

- **Range of spatial correlation:** Vary the range of correlation in the latent spatial random field. For the Matérn class, the spatial range parameter ϕ would be based on the maximum distance between the locations. One example would set ϕ to be 25%, 50%, and 75% of the maximum distance between locations.

- **Signal-to-noise ratio:** For the Matérn class, vary Ratio $= \sigma^2/\tau^2$, the ratio of the partial sill parameter σ^2 and nugget parameter τ^2 of the spatial random field. Higher ratios yield latent processes with less micro-scale variation (less noise), and lower ratios generate fields with higher micro-scale variation (more noise).

- **Sample Size:** Fit SGLMMs to small, moderate, and large samples. This scenario is designed to test the computational efficiency of each basis representation approach.

Using the various basis functions, a spatial generalized linear mixed model (SGLMM) would be fit on each simulated dataset. Metrics of interest include: (1) predictive ability, such as out-of-sample root mean squared prediction error; (2) point and interval estimates for model parameters; (3) model-fitting wall times; and (4) effective samples per second. In addition, it is important to assess the difficulty and computational costs of constructing the spatial basis functions. These are non-trivial fixed costs that may dissuade users from using certain spatial basis functions. For instance, radial basis functions (Cressie and Johannesson, 2008; Katzfuss, 2017; Nychka et al., 2015) require careful selection of knots. Spatial eigenfiltering approaches (Hughes and Haran, 2013; Griffith, 2003) require a costly eigendecomposition performed prior to model fitting, and may require careful rank selection.

5.3.5 Mixed Spatial Basis Functions

Past studies have considered a mixture of basis functions; however, these basis functions typically come from the same class (e.g. radial, wavelets, empirical orthogonal functions, eigenvectors). Examples include Katzfuss (2017) and Nychka et al. (2015), which employ radial basis functions at varying resolutions and knot locations. However, there has not been much research on mixing basis functions across classes.

I propose a variable selection approach to choose the appropriate subset of basis functions from a larger pool of basis functions. The larger pool consists of spatial basis functions from the various classes (see previous point). As pool of basis functions increases, it is computationally prohibitive to examine the 2^N subsets of basis functions, where N is the total size of the pool.

One potential approach would be to apply a variable selection approach like LASSO (Tibshirani, 1994) or use Bayesian variable selection approaches like the spike-and-slab prior distributions Mitchell and Beauchamp (1988); George and McCulloch (1993); Ishwaran et al. (2005) on the basis coefficients. The potential contributions are two-fold. First, the chosen basis functions can be comprised of a mixture of smooth and rough processes; hence, better representing the underlying spatial processes than a single class of basis functions. To illustrate, the resulting

basis functions may consist of spatial patterns (e.g. empirical orthogonal functions, spatial eigenfiltering, or reparameterization approaches) and radial basis functions at finer resolutions. Second, variable selection can automatically determine the rank of the basis functions matrix. This would be a considerable improvement over the automated rank selection heuristic presented in sections 2.4.4 and 3.3.3.

Appendix A

Particle-Based Approach for Computer Model Calibration

A.1 Parameter Descriptions

We calibrate 11 model parameters of the PSU3D-ICE model. The parameter descriptions are as follows:

1. **OCFACMULT:** A dimensionless coefficient multiplying the rate of sub-oceanic melting or freezing calculated at the base of floating ice shelves (Pollard et al., 2016; Pollard and DeConto, 2012a). It corresponds to parameter κ in equation 17 of Pollard and DeConto (2012a). The calculation of sub-ice-shelf melt rate primarily depends on the temperature of nearby oceanic water at 400 m beneath sea level (Pollard and DeConto, 2012a).

2. **OCFACMULTASE:** A dimensionless coefficient that modifies the sub-oceanic ice shelf melting or freezing rate in the Amundsen Sea Embayment of the West Antarctic Ice Sheet (Chang et al., 2016a). Oceanic melting may occur at a different rate here due to stronger regional circulation (Jacobs et al., 2011).

3. **CRHSHELF:** A dimensionless multiplier applied uniformly to basal sliding coefficients for continental shelf areas (modern ocean areas). It multiplies the basal sliding coefficients C' in equation 10 of Pollard and DeConto (2012a),

which have units of m year^{-1} Pa^{-2}.

4. **CRHFAC:** A dimensionless multiplier applied uniformly to basal sliding coefficients for areas with modern grounded ice and was calculated previously using a simple inverse method(Pollard and DeConto, 2012b). It multiplies the basal sliding coefficients C' in equation 10 of Pollard and DeConto (2012a), which have units of m year^{-1} Pa^{-2}.

5. **ENHANCESHEET:** A dimensionless coefficient multiplying the rheologic coefficient in the calculation of the viscous vertical-shearing deformation of ice. This calculation uses the shallow ice approximation (SIA), usually the dominant mode of flow for grounded ice. It corresponds to E in equation 16 of Pollard and DeConto (2012a).

6. **ENHANCESHELF:** A dimensionless coefficient multiplying the rheologic coefficient in the calculation of the viscous horizontal-stretching deformation of ice. This calculation uses the shallow shelf approximation (SSA), usually the dominant mode of flow for floating ice. It corresponds to E in equation 16 of Pollard and DeConto (2012a).

7. **FACEMELTRATE:** A dimensionless coefficient multiplying the melt rate of vertical ice cliffs in contact with warm ocean water at the edges of ice shelve (Pollard and DeConto, 2012a).

8. **TAUASTH:** The e-folding time, for local asthenospheric relaxation towards isostatic equilibrium, in the calculation of bedrock response to varying ice loading and unloading. Units are in years, and it corresponds to τ in equation 33 of Pollard and DeConto (2012a).

9. **CLIFFVMAX:** The maximum erosional retreat rate for unstable marine ice cliffs exceeding approximately 100 meters in sub-aerial height (Pollard et al., 2015). This is the horizontal material velocity of cliff wastage into the upstream solid ice, in the parameterization of marine ice cliff instability (MICI). Units are in meters per year.

10. **CALVLIQ:** Scaling depth for the deepening of surface crevasses by hydrofracturing due to surface melt and rainfall. Its units are meters of crevasse

depth and is the crevasse deepening produced by a surface melt plus rainfall rate of 1 meter per year. It corresponds to the constant 100 in equation B.6 of Pollard and DeConto (2012a).

11. **CALVNICK :** A dimensionless coefficient multiplying the combined total depth of crevasses in the calving parameterization. This depth is compared to the actual ice-shelf thickness in the model's calving parameterization (Pollard et al., 2015; Nick et al., 2010). It multiplies the parameter r in equation B.7 of Pollard and DeConto (2012a).

A.2 Simulated Example

We provide additional details pertaining to the simulation study using $N = 2000$ particles from Section 5. Maps of the model outputs and observations are provided in Figure A.2. The simulated calibration experiment went through four sampling-importance-resampling cycles with corresponding incorporation increments $\gamma = \{0.100, 0.15, 0.27, 0.47\}$. Our adaptive likelihood incorporation schedule chose four sampling-importance-resampling cycles. In the first cycle, our algorithm chose a incorporation increment $\gamma_1 = 0.1$, which yields an effective sample size (ESS) of 169.5. In the second cycle, the algorithm chose an incorporation increment $\gamma_2 = 0.15$ with a corresponding ESS of 1000. For the third cycle, the selected incorporation increment is $\gamma_3 = 0.27$ with a corresponding ESS of 1000. In the fourth and final cycle, we use an incorporation increment of $\gamma_4 = 0.47$ with a corresponding ESS of 1143. Figure A.3 shows the chosen incorporation increments and corresponding ESS for each cycle. Figure A.4 displays posterior parameter densities after each cycle.

In the mutation stage, we chose the baseline number of Metropolis-Hastings updates to be 10 updates. Our algorithm determined that the empirical distribution of our stopping metric, model parameter θ, stabilizes after 20 total iterations. The stopping criterion is met once the Batthacharyya distance of the empirical samples at the 20-th mutation update and the 10-th update is less than a pre-determined threshold.

A.3 Emulation-Calibration Details

We provide additional details regarding the comparative analysis performed in Section 5.3. Available paleoclimate and observational data include the Antarctic ice sheet's contribution to sea level change in the (1) Pliocene era; (2) Last Interglacial Age; (3) Last Glacial Maximum; (4) the total volume of the Antarctic ice sheet in the modern era; and (5) total grounded area of the Antarctic ice sheet in the modern era. For the comparison study, we omit the binary observations (ice vs. no ice) obtained at the 10 strategic locations (Manuscript Section 6.1). We use the same prior distributions for our model parameters as provided in Section 6.1 of the manuscript.

For the three-parameter emulation-calibration example, we select OCFAC-MULT, CALVLIQ, and CLIFFVMAX as the calibration parameters and fix the remaining eight parameters. To train the Gaussian process emulator, we use PSU3D-ICE output obtained at 512 different input parameter settings. We generate the input parameter settings using a full factorial design, which includes eight discrete levels for each model parameter. The eight levels span the uniform prior distribution ranges as provided in Section 6.1 of the manuscript. We fit a separate Gaussian process for each modern and paleo-climate observational record (5 total); in addition, we fit a Gaussian process for the Antarctic ice sheet contribution to sea level change in 2100, 2200, 2300, 2400, and 2500. Each Gaussian process has the form $Y(\theta) \sim \mathcal{GP}(\mu(\theta; \beta_0; \beta), C(\theta, \theta'; \sigma^2, \phi))$, where the mean function $\mu(\theta; \beta_0; \beta) = \beta_0 + \beta\theta$ includes an intercept and a linear trend. We use a squared exponential covariance function, $C(\theta, \theta'; \sigma^2, \phi) = \sigma^2 \prod_{i=1}^{p} \exp\{-\frac{(\theta_i - \theta_i')^2}{\phi_i}\}$, where $\theta \in \mathbb{R}^p$ $\phi = (\phi_1, ..., \phi_p)$. We estimate the Gaussian process parameters, (β, σ^2, ϕ), through maximum likelihood estimation. we fit the Gaussian process emulator using the **mlegp** R package (Dancik and Dorman, 2008). The 3-parameter Gaussian process emulator has a low out-of-sample cross validated root mean squared prediction error as shown in Table A.3.

In the 11-parameter emulation-calibration study, we implement a two-part emulation-calibration method using all model parameters. We run the PSU3D-ICE model at 512 input parameter settings chosen through a Latin Hypercube Design (LHC). The LHC samples span the ranges of the prior distributions pro-

vided in Section 6.1 of the manuscript. Similar to the three-parameter case, we fit a Gaussian process emulator via maximum likelihood estimation. The 11-parameter Gaussian process emulator has a high out-of-sample cross validated root mean squared prediction error, as shown in Table A.3. This can be attributed to the low-fidelity emulator trained using a small number of design points (512) to explore an 11-dimensional parameter space.

	3 Parameter Emulator RMSE	11 Parameter Emulator RMSE
Pliocene	0.20	1.08
Last Interglacial	0.15	0.87
Last Glacial Maximum	0.02	6.18
Modern SLE	0.25	7.02
Modern Volume	0.18	3.73
Year 2100	0.27	5.71
Year 2200	0.37	6.40
Year 2300	0.23	1.92
Year 2400	0.26	0.87
Year 2500	0.23	0.82

Table A.1. Out-of-sample cross validated root mean squared prediction error (RMSE) for a Gaussian process emulator with 3 parameters and 11 parameters. The three-parameter emulator exhibits low RMSE across all observations and projections. The 11-parameter emulator has a high RMSE, which is indicative of a low-fidelity, or inaccurate, surrogate model.

A.4 Prior Sensitivity Analysis

We conduct a prior sensitivity analysis using two sets of prior distributions provided by domain experts. The first set of prior distributions are from the main calibration experiment in Section 6.1 of the manuscript. The second set of prior distributions includes extended ranges for the model parameters. Note that we change the prior distribution for model parameters – CALVNICK, TAUASTH, CALVLIQ, CLIFFVMAX, FACEMELTRATE – from a uniform distribution to a log-uniform distribution. The second set of prior distributions are:

- $\log_{10}(\theta_{OCFACMULT}) \sim \mathcal{U}(-2, 2)$
- $\log_{10}(\theta_{OCFACMULTASE}) \sim \mathcal{U}(-1.5, 2.5)$

- $\log_{10}(\theta_{CALVNICK}) \sim \mathcal{U}(-2, 2)$
- $\log_{10}(\theta_{CRHSHELF}) \sim \mathcal{U}(-9.5, -1.5)$
- $\log_{10}(\theta_{CALVLIQ}) \sim \mathcal{U}(1, 3)$
- $\log_{10}(\theta_{FACEMELTRATE}) \sim \mathcal{U}(-1, 3)$
- $\log_{10}(\theta_{ENHANCESHEET}) \sim \mathcal{U}(-2, 2)$
- $\log_{10}(\theta_{CRHFAC}) \sim \mathcal{U}(-3, 3)$
- $\log_{3}(\theta_{TAUASTH}) \sim \mathcal{U}(2, 4)$
- $\log_{6}(\theta_{CLIFFVMAX}) \sim \mathcal{U}(0, 5)$
- $\log_{0.3}(\theta_{ENHANCESHELF}) \sim \mathcal{U}(-2, 2)$

A.5 Toy Example Comparative Study

One major contribution of this study is reducing the number of sequential likelihood evaluations. Each likelihood evaluation requires a computer model runs, which are the dominant costs of our approach. We introduce an adaptive likelihood incorporation schedule which is automated. In a standard implementation of the particle-based method, we must set the total number of sampling-importance-resampling cycles and the total number of mutation runs per cycle. Here, we compare results from a standard implementation to those using our fast adaptive method. In the standard implementation, we set the total number of cycles to be 6 and the total number of Metropolis-Hastings updates (for the mutation stage) to be 45. These chosen values are based on the available computing resources, namely a 12-hour walltime limit for each mutation cycle. In this comparison study, we use the five modern and paleoclimate records as observations; spatial constraints are omitted.

Upon examining the standard implementation, we observe that the distribution of the particles do not change after 10 Metropolis-Hastings updates of the mutation stage. Therefore, the remaining 35 Metropolis-Hastings updates are redundant. Moreover, the posterior densities of the model parameters (Figure A.8),

observational records (Figure A.9), and sea level projections (Figure A.10) for both methods (standard vs. adaptive) are very much similar.

A.6 Fundamental Equations for the PSU3D-ICE Model

In the main paper, the ice-sheet model is treated as a 'black box' within our calibration framework. To provide an overall picture of the physical ice-sheet model, we presents its main equations. In a sense, they are its most fundamental partial and ordinary differential equations used to time-step the state of the ice sheet forward in time. Other equations, mostly parameterizations of local processes, are also used but are not as fundamental in the sense mentioned above.

The basic aspects of continental ice sheets and models are as follows. Ice cover on continental scales forms a dome, several kilometers thick in central regions and sloping downward to its margins at much lower elevations. Thickening due to annual snowfall (which compacts to ice) in interior regions is balanced by ice velocities towards the margins, as the ice deforms slowly under its own weight. Ice is lost mostly near the margins by surface melt, basal melt, oceanic melt, and calving of marginal vertical ice faces. If the ice reaches the ocean, it can flow across the grounding line (where the bed is below sea level and ice becomes afloat), and form floating ice shelves with thicknesses of 100's m and extents of 100's km. Horizontal ice velocities are ∼1 to 10 meters/years in much of the central interior, increasing to ∼100's to ∼1000 meters/year in marginal ice streams and shelves (Rignot et al., 2011).

Numerical ice-sheet models predict the time-evolving ice thicknesses and temperature distributions, changing due to velocity advection and the local accumulation and ablation processes mentioned above. Ice flow is treated as a non-linear viscous fluid using scaled (simplified) equations, separately for horizontal stretching and for the vertical shear of horizontal velocities. Slow depression and rebound of the bedrock beneath the changing ice load is also modeled, as this affects ice surface elevations and ocean depths at grounding lines. These basic aspects are common to many large-scale ice-sheet models, and are described in detail in Pol-

lard and DeConto (2012a) and Pollard et al. (2015).

I. Ice Thickness

$$\frac{\partial h}{\partial t} + \frac{\partial(\bar{u}h)}{\partial x} + \frac{\partial(\bar{v}h)}{\partial y} = \text{SMB} - \text{BMB} - \text{OMB} - \text{CMB} - \text{FMB},$$

where h is ice thickness, $\bar{u}$ is the mean horizontal ice velocity in the x direction, $\bar{v}$ is the mean horizontal ice velocity in the y direction, SMB is the surface mass balance, BMB is the basal melting (if grounded), OMB is the oceanic sub-ice melting or freezing (if floating), CMB is the calving loss (floating edge), and FMB is the face melt loss (floating or tidewater vertical face).

II. Velocity Stretching:

$$\frac{\partial}{\partial x}\Big[\frac{h}{A\sigma^{n-1}}\Big(2\frac{\partial \bar{u}}{\partial x}+\frac{\partial \bar{v}}{\partial y}\Big)\Big]+\frac{\partial}{\partial y}\Big[\frac{h}{2A\sigma^{n-1}}\Big(\frac{\partial \bar{u}}{\partial y}+\frac{\partial \bar{v}}{\partial x}\Big)\Big] = \rho_i g h\frac{\partial h_s}{\partial x}+\frac{1}{C'^{1/m}}\frac{1}{|u_b|^{1-\frac{1}{m}}}u_b, \tag{A.1}$$

$$\frac{\partial}{\partial y}\Big[\frac{h}{A\sigma^{n-1}}\Big(2\frac{\partial \bar{v}}{\partial y}+\frac{\partial \bar{u}}{\partial x}\Big)\Big]+\frac{\partial}{\partial x}\Big[\frac{h}{2A\sigma^{n-1}}\Big(\frac{\partial \bar{u}}{\partial y}+\frac{\partial \bar{v}}{\partial x}\Big)\Big] = \rho_i g h\frac{\partial h_s}{\partial y}+\frac{1}{C'^{1/m}}\frac{1}{|v_b|^{1-\frac{1}{m}}}v_b, \tag{A.2}$$

where $\bar{u} = \bar{u}_i + u_b$ and $\bar{v} = \bar{v}_i + v_b$. Here, $\bar{u}_i$ is mean horizontal velocity from vertical shearing, and u_b is basal sliding velocity in the x direction. Similarly, $\bar{v}_i$ is mean horizontal velocity from vertical shearing, and v_b is basal sliding velocity in the y direction. u_i is the horizontal velocity in the x direction from vertical shearing (i.e., minus its value at the base), and v_i is the horizontal velocity in the y direction from vertical shearing. A is the ice rheological coefficient, σ is the effective stress (second invariant of the stress tensor), $n = 3$ is the ice rheological exponent, and g is gravitational acceleration. C' is the basal sliding coefficient between bed and ice and m is the basal sliding exponent. h_s is ice surface elevation, where

$$h_s = \begin{cases} h + h_b & \text{, if grounded} \\ (\frac{p_w - p_i}{p_i})h & \text{, if floating,} \end{cases}$$

where h_b is the bedrock elevation, ρ_w is the ocean water density, and ρ_i is ice density.

III. Velocity Shearing:

$$\frac{\partial u_i}{\partial z} = -2A\sigma^{n-1}\Big(\rho_i g h \frac{\partial h_s}{\partial x} - L_x\Big) \times \Big(\frac{h_s - z}{h}\Big),$$
$$\frac{\partial v_i}{\partial z} = -2A\sigma^{n-1}\Big(\rho_i g h \frac{\partial h_s}{\partial y} - L_y\Big) \times \Big(\frac{h_s - z}{h}\Big),$$

where z is the vertical elevation, L_x is the left hand side of Equation A.1, and L_y is the left hand side of Equation A.2.

IV. Temperature:

The prognostic equation for internal ice temperatures T is

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + w\frac{\partial T}{\partial z} = \frac{1}{\rho_i c_i}\frac{\partial}{\partial z}\Big(\frac{k\partial T}{\partial z}\Big) + \frac{Q}{\rho_i c_i},$$

where $u = u_b + u_i(z)$, $v = v_b + v_i(z)$, and w is deduced from continuity. k is the ice thermal conductivity, Q is internal deformational heating, and c_i is the specific heat of ice.

V. Bedrock Elevation:

The rate of change of bedrock elevation is given by:

$$\frac{\partial h_b}{\partial t} = -\frac{1}{\tau}(h_b - h_b^{eq} + w_b),$$

where h_b^{eq} is its equilibrium value and $\tau = 3000$ years is the asthenospheric isostatic relation time scale. The downward deflection of the fully relaxed response (as if the asthenosphere had no lag), w_b, is given by:

$$D\nabla^4 w_b + \rho_b g w_b = q,$$

where D is the flexural rigidity of the lithosphere, ρ_b is the bedrock (asthenospheric) density, and the applied load q is:

$$q = \rho_i g(h - h^{eq}) + \rho_w g(h_w - h_w^{eq}),$$

where h_w is ocean column thickness, h_w^{eq} is ocean column thickness in the equilib-

rium state, and h^{eq} is ice thickness in the equilibrium state.

	Months
OCFACMULT	15.3
OCFACMULTASE	15.0
CALVNICK	15.2
CRHSHELF	15.1
TAUASTH	16.3
CALVLIQ	14.6
CLIFFVMAX	18.5
FACEMELTRATE	14.6
ENHANCESHEET	13.8
ENHANCESHELF	12.6
crhfac	16.5

Table A.2. Estimated time to obtain the desired effective sample size of 1533 using the all-at-once random walk Metropolis-Hastings algorithm. Note that the particle-based approach utilized 2015 particles with an ESS of 1533.

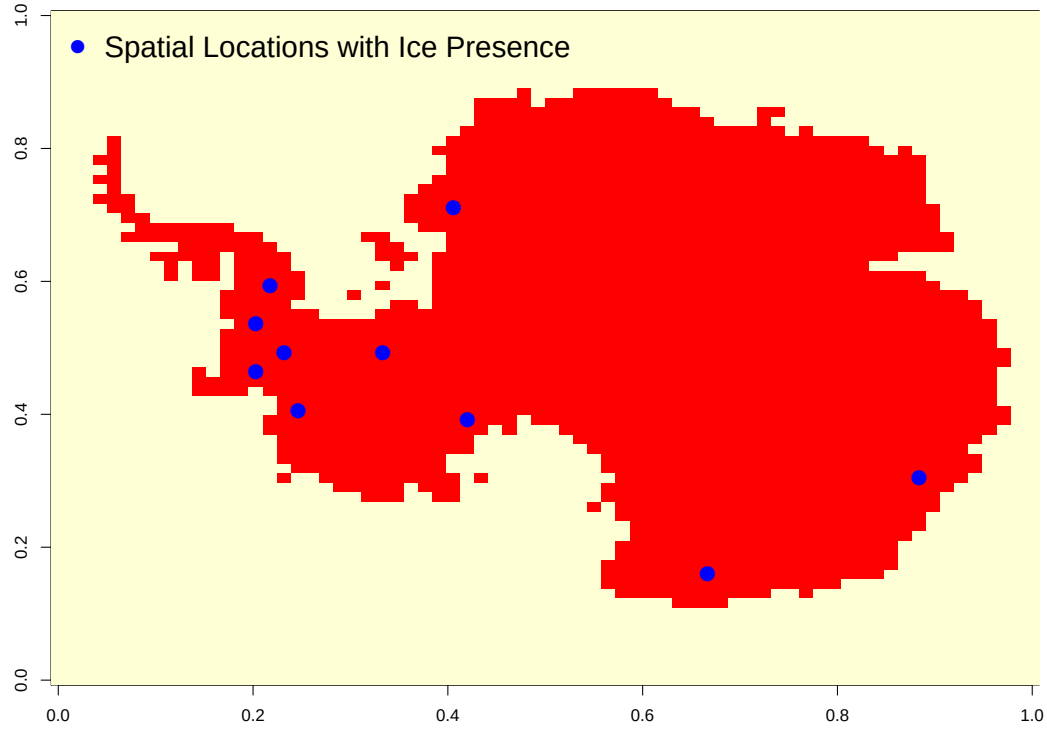

Figure A.1. Modern observations of ice presence obtained via the Bedmap2 project. The blue dots indicate locations where there is confirmed ice presence.

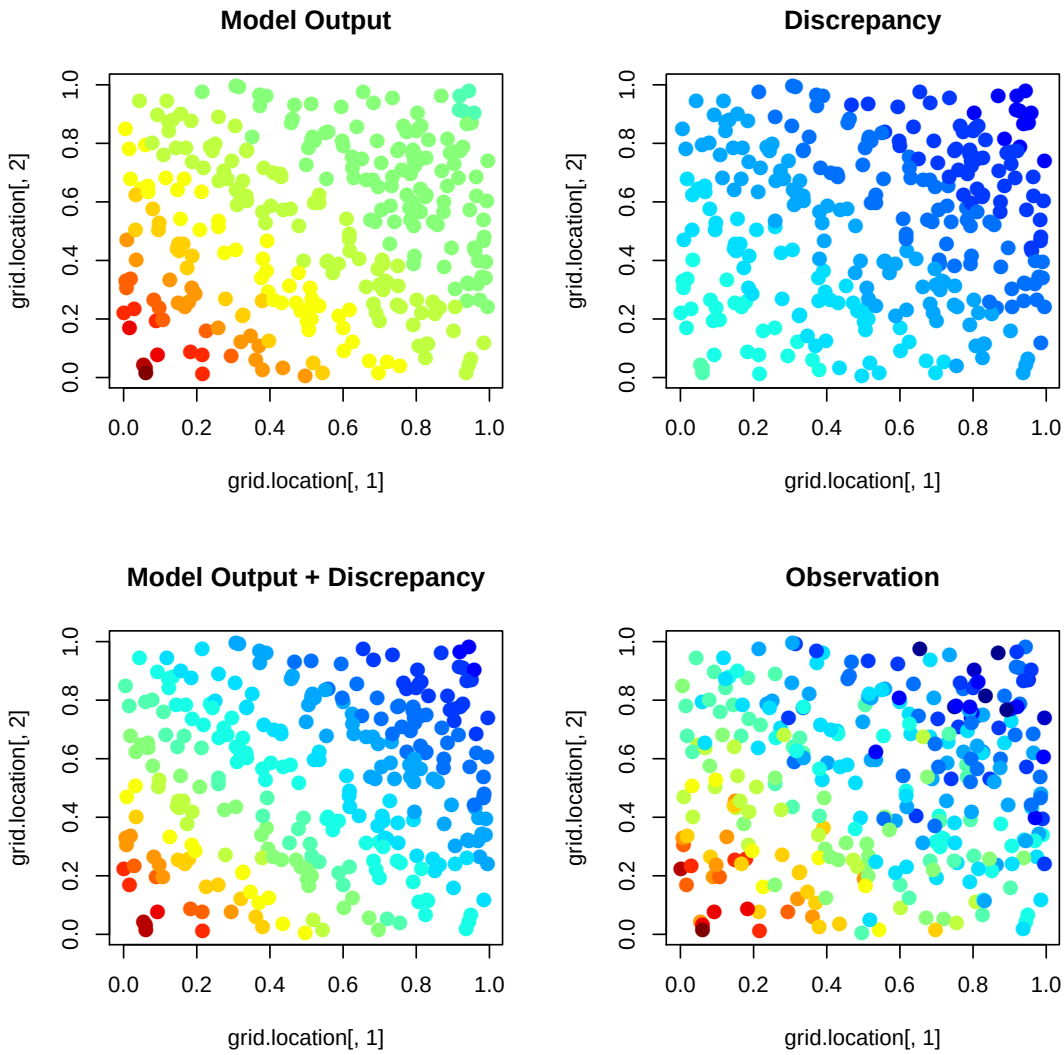

Figure A.2. (Top left) Map of the model output from the toy example. (Top right) Map of the systematic and also spatially correlated data-model discrepancy. (Bottom left) Map of the sum of the model output and discrepancy. (Bottom right) Map of the observations, which is the sum of the model output, discrepancy, and iid observational error.

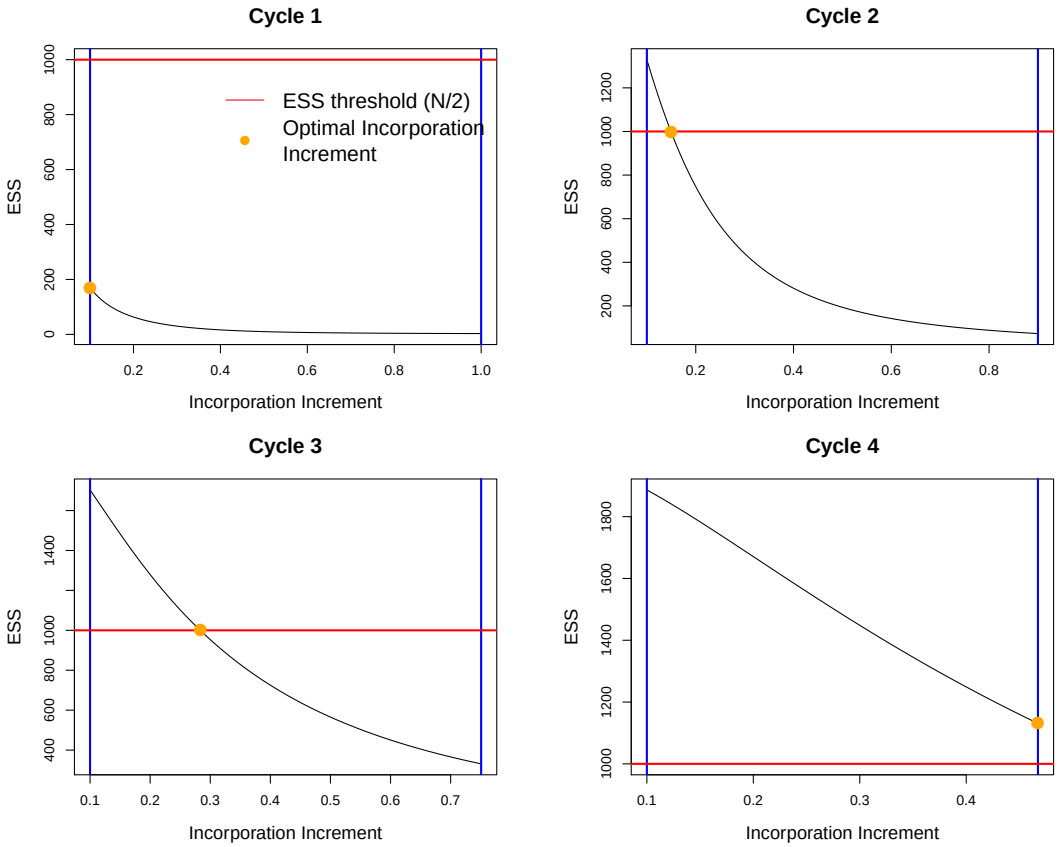

Figure A.3. Incorporation increment γ_t selection for the simulated example. Each panel corresponds to a cycle (4 total). The x-axis denotes possible values for the incorporation increment γ_t and the y-axis denotes the corresponding effective sample size (ESS). The red line represents the ESS threshold set at $N/2$. The orange point denotes the optimal incorporation increment and the corresponding ESS at each cycle.

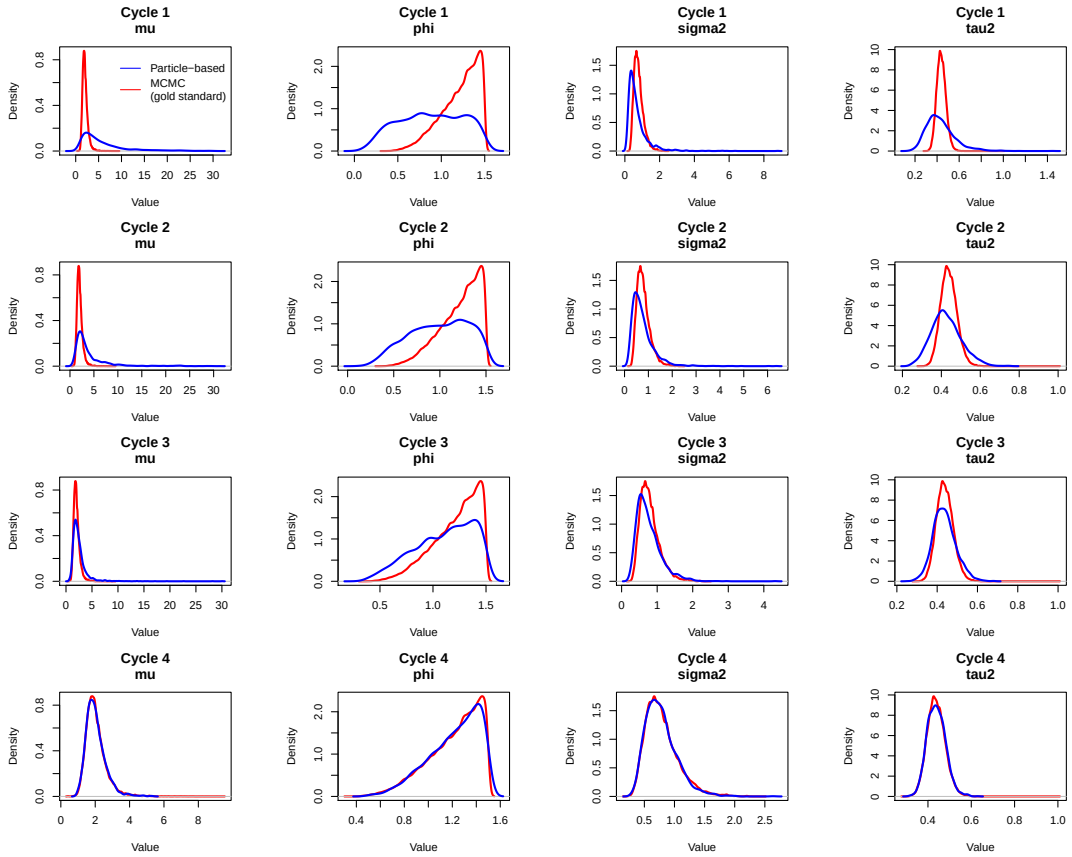

Figure A.4. Posterior densities for the simulated example after each cycle. Each row corresponds to a cycle, and each column corresponds to a model parameter. The blue lines represent the density of the posterior samples from the particle-based approach, and the red lines denote the density of the posterior samples obtained form MCMC (gold standard). Note that the particle-based approach provides a good approximation to the MCMC-based approach. However, the particle-based approach requires just 80 model evaluations as opposed to 100k for the MCMC-based approach.

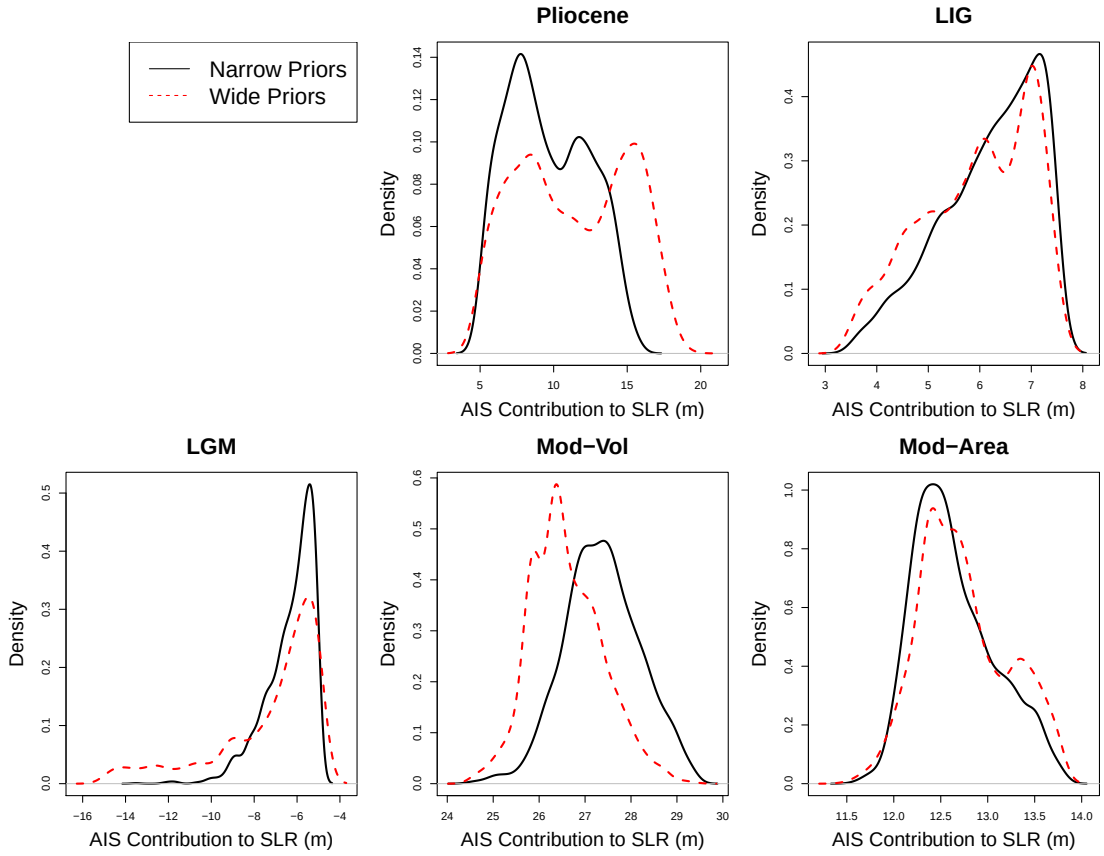

Figure A.5. Posterior densities of observational records using expert prior distributions (solid black lines) and wider expert prior distributions (dashed red lines). Wider expert priors result in a bi-modal distribution for the AIS contribution to sea level rise in the Pliocene and lower modern volume, both in point estimate and 95% credible intervals.

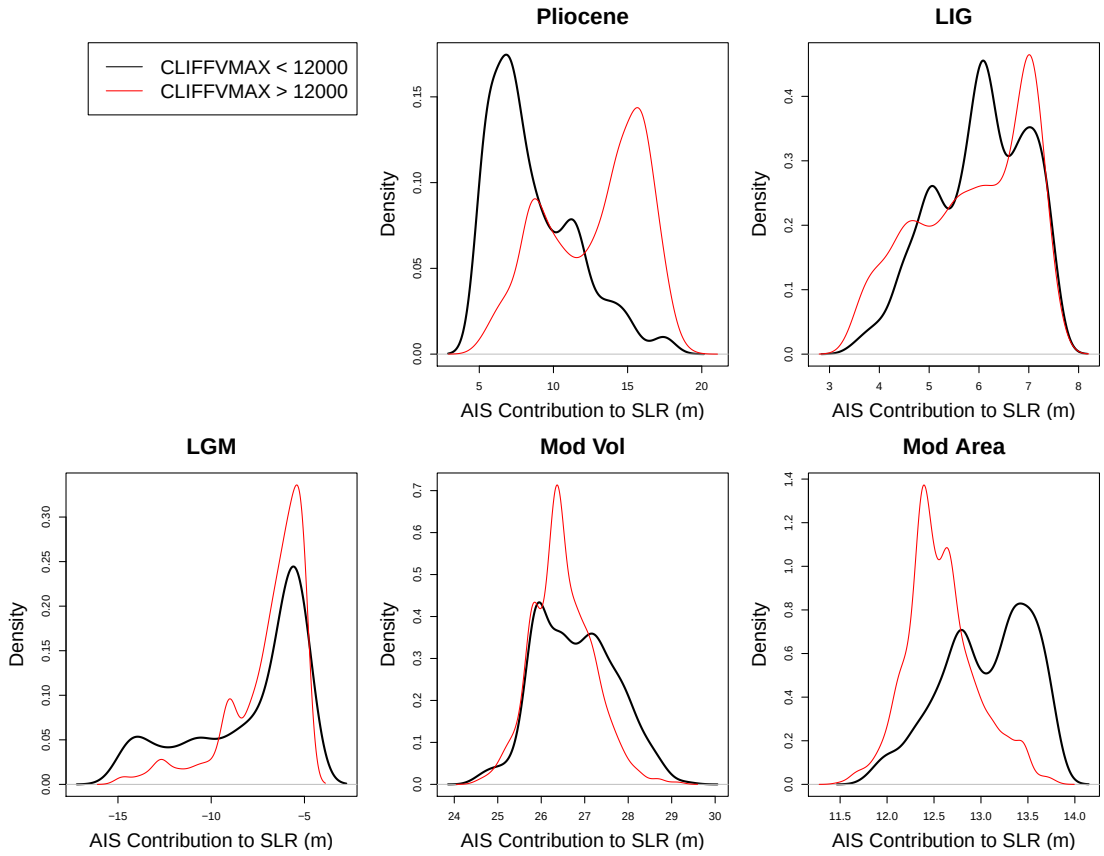

Figure A.6. Posterior densities of observational records using the wider expert prior distributions. The posterior densities are split for values of CLIFFVMAX less than 12 km per year (black lines) and greater than 12 km per year (red lines). Higher values of CLIFFVMAX results in higher values (point estimates and 95% credible intervals) of the Antarctic ice sheet's contribution to sea level rise in the Pliocene and lower modern volume.

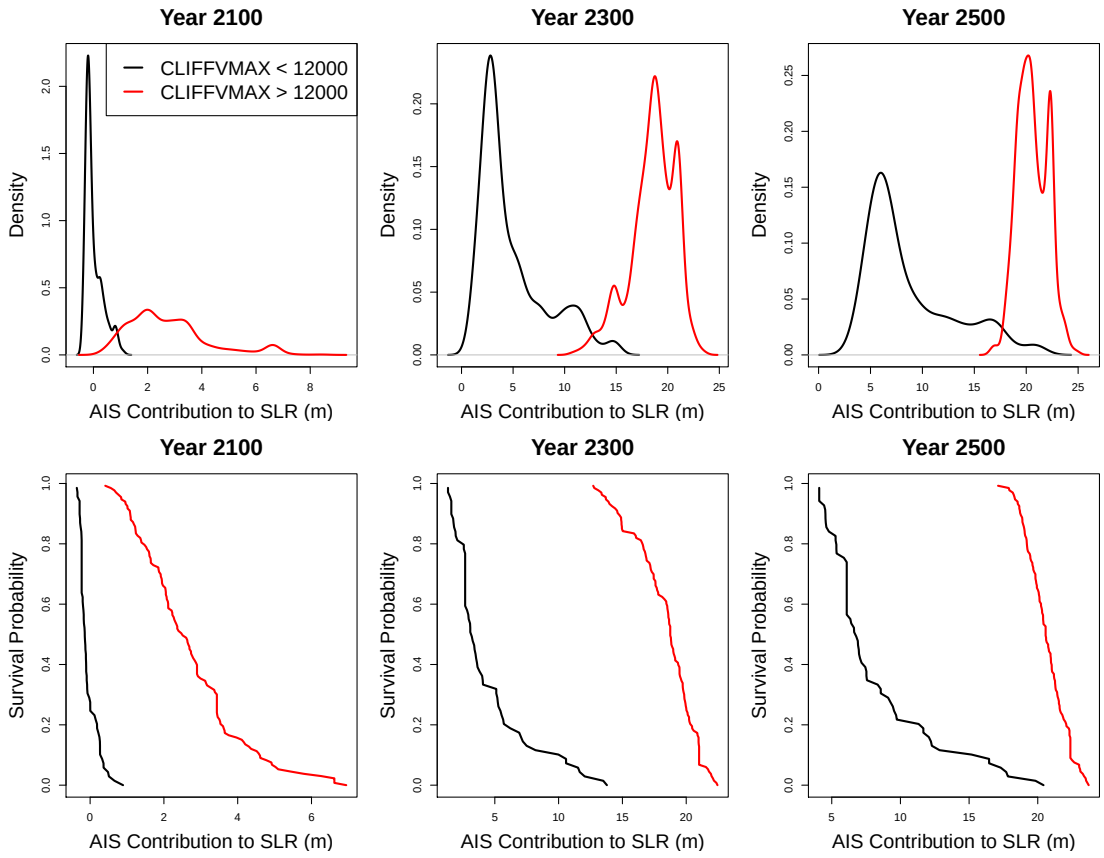

Figure A.7. (Top Panel) Posterior densities of the projected Antarctic ice sheet's contribution to sea level change in 2100, 2300, and 2500 using the wider expert prior distributions. The posterior densities are split for values of CLIFFVMAX less than 12 km per year (black lines) and greater than 12 km per year (red lines). (Bottom Panel) Empirical survival function of the projected Antarctic ice sheet's contribution to sea level change in 2100, 2300, and 2500 for higher CLIFFVMAX values (solid black lines) and lower CLIFFVMAX values (red lines). Larger values of CLIFFVMAX results in considerably higher projections of future sea level rise, both in point estimates and 95% credible intervals.

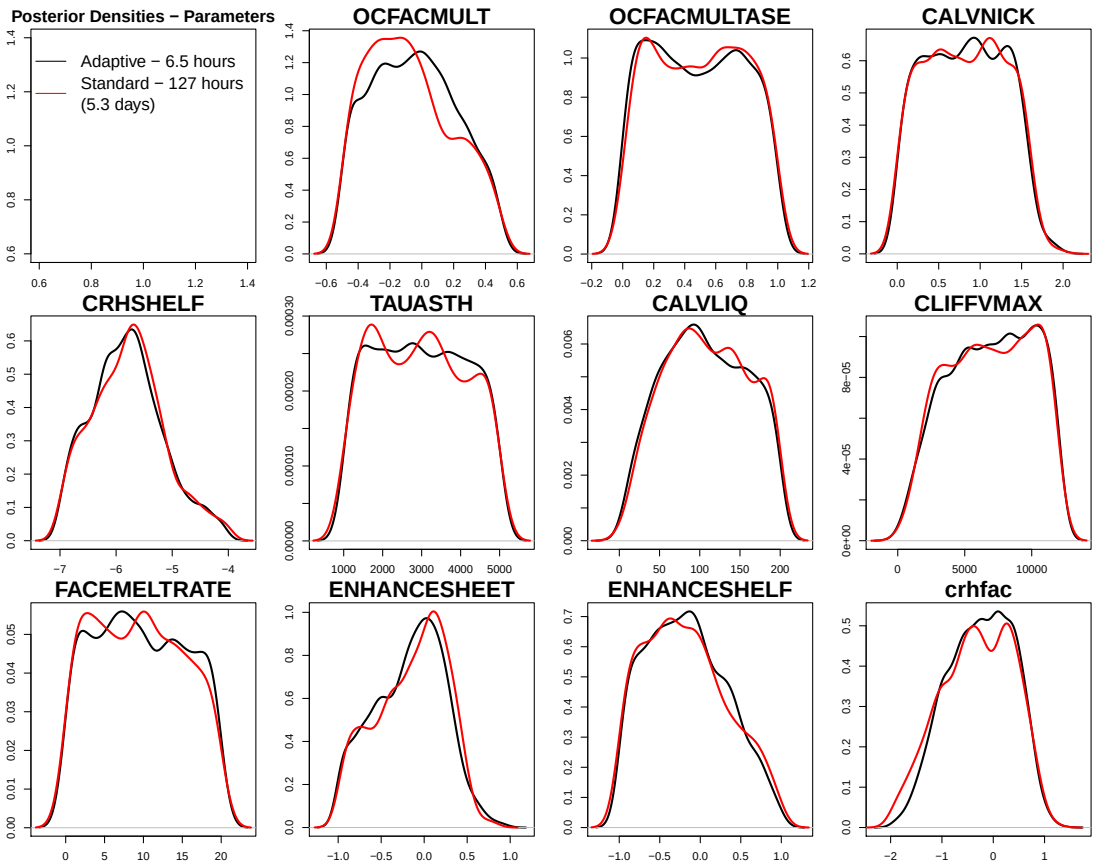

Figure A.8. Posterior densities of model parameters using adaptive particle-based approach (solid black lines) and the standard particle-based approach (dashed red lines). The adaptive particle-based approach goes through 4 cycles and runs 14 updates in the mutation stage with a total calibration wall time of 6.5 hours. The standard particle-based approach goes through 10 cycles and runs 45 updates in the mutation stage with a total calibration wall time of 127 hours (5.3 days). Posterior densities for both methods are comparable.

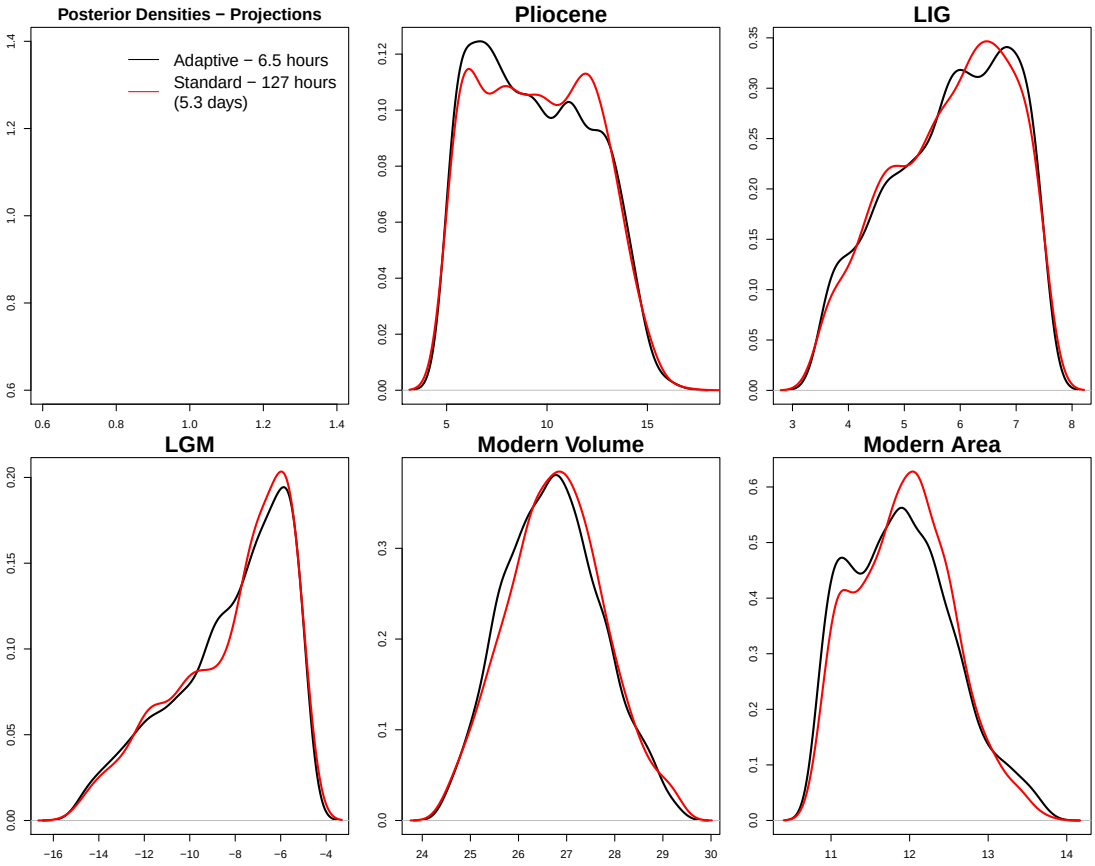

Figure A.9. Posterior densities of observations using the adaptive particle-based approach (solid black lines) and the standard approach (dashed red lines). The adaptive particle-based approach goes through 4 cycles and runs 14 updates in the mutation stage with a total calibration wall time of 6.5 hours. The standard particle-based approach goes through 10 cycles and runs 45 updates in the mutation stage with a total calibration wall time of 127 hours (5.3 days). Posterior densities for both methods are comparable.

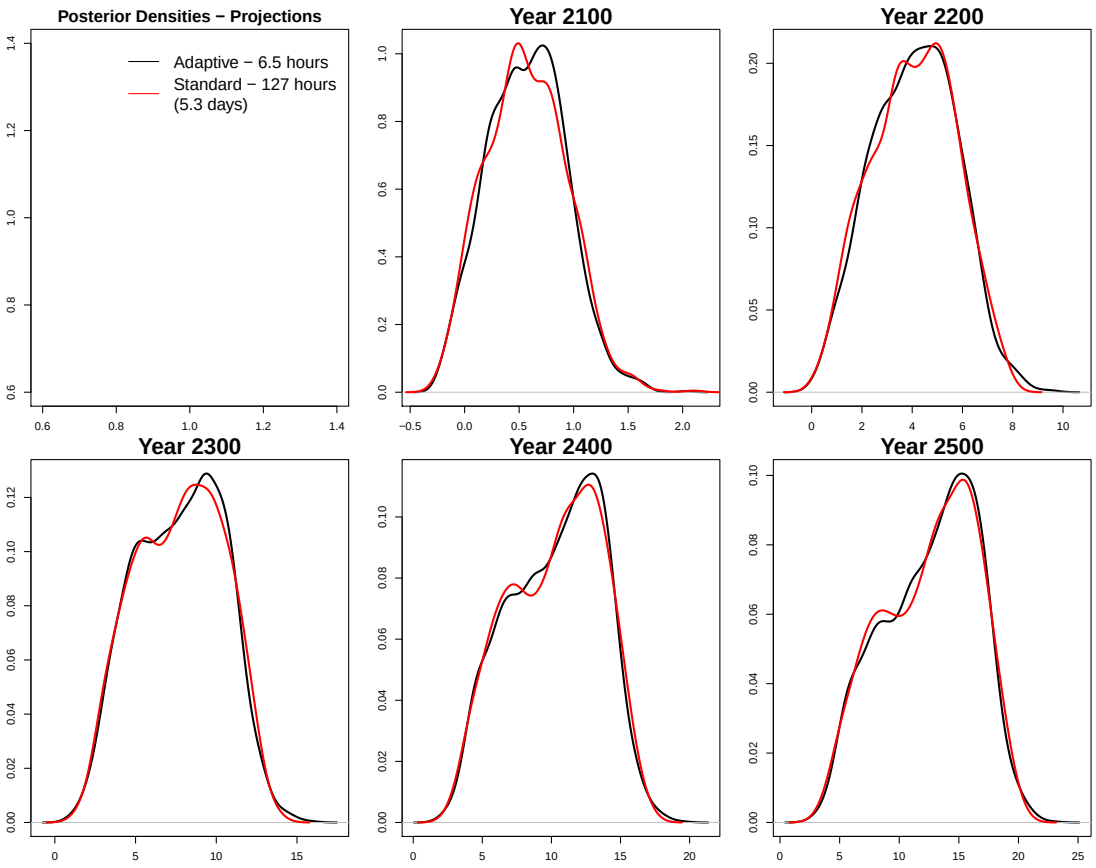

Figure A.10. Posterior densities of projections using adaptive particle-based approach (solid black lines) and the standard approach (dashed red lines). The adaptive particle-based approach goes through 4 cycles and runs 14 updates in the mutation stage with a total calibration wall time of 6.5 hours. The standard particle-based approach goes through 10 cycles and runs 45 updates in the mutation stage with a total calibration wall time of 127 hours (5.3 days). Posterior densities for both methods are comparable.

Appendix B

PICAR: Projection Intrinsic Conditional AutoRegression

B.1 Examples of Hierarchical Spatial Models

Here we provide two examples of hierarchical spatial models. The first is a spatially varying coefficient model and the second is a cumulative-logit model for ordered categorical data.

Spatially Varying Coefficient Models

Spatially varying coefficient models (Gelfand et al., 2003) consider cases where the fixed effects β vary across space. For the case with a single predictor $X(s)$, the data model is $Z(s) = \beta_0 + \beta_1 X(s) + \beta_1(s)X(s) + w(s) + \epsilon(s)$, where β_0 is the intercept, β_1 is the fixed effect, $\beta_1(s)$ is the spatially varying coefficient term, and $w(s)$ and $\epsilon(s)$ are the spatial random effects and micro-scale measurement errors, respectively. Here, $\mathbf{B} = (\beta_1(s_1), ..., \beta_1(s_n))$ is the n-dimensional vector of spatially varying coefficients, and $\mathbf{B} \sim N(0, \sigma^2_\beta R_{\phi_\beta})$ where σ^2_β is the partial sill and ϕ_β is the range parameter for the spatial random process $\mathbf{B}$.

For cases with k predictors, we have the following hierarchical spatial model:

$$
\begin{aligned}
&\textbf{Data Model:} && Z(s)|\eta(s) \sim f(\eta(s)) \\
& && \eta(s) = X(s)\beta + X(s)\beta(s) + w(s) + \epsilon(s) \\
&\textbf{Process Model:} && (\mathbf{W}, \mathbf{B})^T|\phi, \mathbf{T} \sim \mathcal{N}(\mathbf{0}, R_\phi \otimes \mathbf{T}) \\
& && \epsilon(s)|\tau^2 \sim N(\mathbf{0}, \tau^2) \\
&\textbf{Parameter Model:} && \beta \sim \pi(\beta), \quad \tau^2 \sim \pi(\tau^2), \quad \phi \sim \pi(\phi), \quad \mathbf{T} \sim \pi(\mathbf{T})
\end{aligned}
$$

where β is the k-dimensional vector of the fixed effects, $\beta(s) = (\beta_1(s), ..., \beta_k(s))$ is a k-dimensional vector of the spatially varying coefficients for location s, $\mathbf{B} = (\beta(s_1), ..., \beta(s_n))$ is the nk-dimensional vector of all spatially varying coefficients, $\mathbf{W} = (W(s_1), ..., W(s_n))$ is the n-dimensional vector of the spatial random effects, R_ϕ and τ^2 are the correlation matrix and nugget variance, and $\mathbf{T}$ is a $(k+1)\times(k+1)$ positive definite matrix.

Cumulative-Logit Models for Ordinal Spatial Data

Ordered categorical (ordinal) data are categorical responses with a natural ordering, and commonly used in survey questionnaires, patient responses in clinical trials, and quality assurance ratings for industrial processes. (Higgs and Hoeting, 2010; Schliep and Hoeting, 2013) develop a hierarchical spatial model for ordinal data. In this study, we examine the proportional-odds cumulative logit model (Agresti, 2010) for ordered categorical data. Let $Z(s)$ be the observations at location $s \in \mathcal{D}$ with J ordered categories. Note that each ordered category corresponds to a probability $\pi(s) = \{\pi_1(s), \pi_2(s), ..., \pi_J(s)\}$, where $\pi_i(s) = \Pr(Z(s) = i)$ for $i = 1, ..., J$. Here, we consider $J - 1$ cumulative probabilities denoted as $\gamma_j(s) = P(Z(s) \leq j) = \pi_1(s) + ... + \pi_j(s)$. The cumulative logit is defined as:

$$\log\left(\frac{P(Z(s) \leq j)}{1 - P(Z(s) \leq j)}\right) = \log\left(\frac{\gamma_j(s)}{1 - \gamma_j(s)}\right) = \theta_j - X(s)\beta - w(s) - \epsilon(s),$$

where θ_j is the intercept or "cutoff" for the j-th category, $X(s)$, β, $w(s)$ and $\epsilon(s)$ are the spatial random effects and micro-scale measurement errors. The model for

the cumulative probabilities γ_j is:

$$\gamma_j(s) = P(Z(s) \leq j) = \frac{\exp\{\theta_j - (X(s)\beta + w(s) + \epsilon(s))\}}{1 + \exp\{\theta_j - (X(s)\beta + w(s) + \epsilon(s))\}}.$$

Consequently, the probabilities for the individual J categories are:

$$P(Z(s) = j) = \begin{cases} \gamma_1(s), & j = 1 \\ \gamma_j(s) - \gamma_{j-1}(s), & 2 \leq j \leq J-1 \\ 1 - \gamma_{J-1}(s), & j = J \end{cases}$$

To avoid identifiability issues, we typically fix the first cutoff to be $\theta_1 = 0$ (Johnson and Albert, 2006). Note that the θ_j's are constrained by the ordering $\theta_j > \theta_k$ for $j > k$. Through a transformation (Higgs and Hoeting, 2010; Albert and Chib, 1997), we can generate unconstrained cutoff parameters $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{J-1})$, where $\alpha_1 = -\infty$, $\alpha_2 = \log(\theta_2)$, and $\alpha_j = \log(\theta_j - \theta_{j-1})$ for $j = 3, ..., J-1$. The inverse transformation is $\theta_j = \sum_{i=1}^{J-1} \exp\{\alpha_i\}$. The hierarchical spatial model framework is as follows:

Data Model:
$$Z(s)|\gamma(s) \sim f(\gamma(s))$$
$$\gamma_j(s)|\beta, \theta, \mathbf{W}, \epsilon(s) = \frac{\exp\{\theta_j - (X(s)\beta + w(s) + \epsilon(s))\}}{1 + \exp\{\theta_j - (X(s)\beta + w(s) + \epsilon(s))\}}$$
$$\theta_j|\alpha = \sum_{i=1}^{J-1} \exp\{\alpha_i\}$$

Process Model:
$$\mathbf{W}|\phi, \sigma^2 \sim N(\mathbf{0}, \sigma^2 R_\phi)$$
$$\epsilon(s)|\tau^2 \sim N(\mathbf{0}, \tau^2)$$

Parameter Model:
$$\alpha \sim p(\alpha), \quad \beta \sim p(\beta), \quad \phi \sim p(\phi), \quad \sigma^2 \sim p(\sigma^2), \quad \tau^2 \sim p(\tau^2)$$

	β_1	β_2
Independent	0.88	0.93
ICAR	0.89	0.91
CAR	0.88	0.95

Table B.1. Binary data simulation study: Coverage probabilities for 100 simulated samples. Columns correspond to the regression coefficients. Rows correspond to the type of precision matrix.

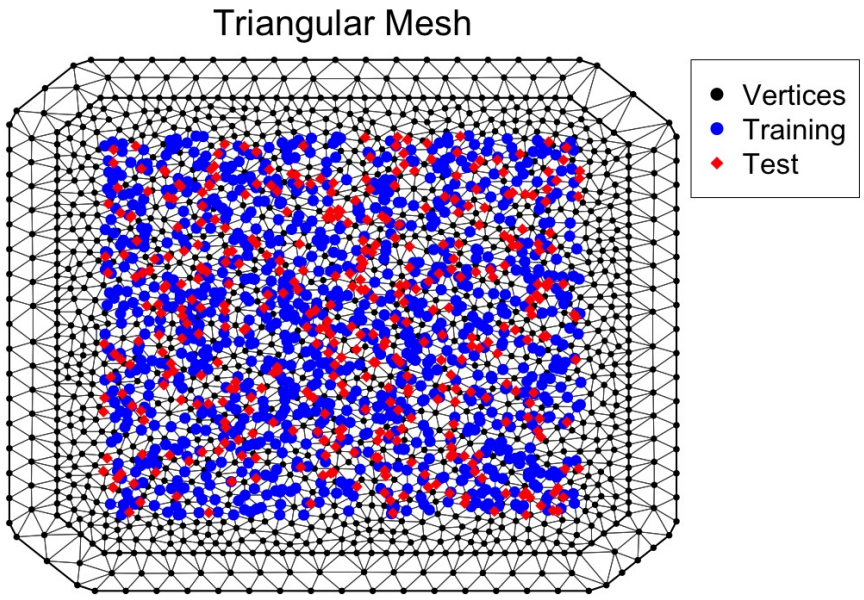

Figure B.1. Triangular Mesh for data in simulation studies. Black points denote the vertices, or nodes, of the triangular mesh. Blue points represent the observation locations used to fit the hierarchical spatial models, and the red points denote the observations locations for the validation sample.

B.2 Simulation study with spatial count observations

We conduct a simulation study using spatial count observations using 100 samples. The regression coefficients and the latent spatial random field are generated in the same way as the binary case. The observations come from a spatial generalized linear mixed model (SGLMM) with a Poisson data model and a log link function. Mesh construction and model fitting details follow closely to the binary case. We select one sample (from the 100 generated samples) as the dataset for the comparative analysis. When comparing across ranks, we elect to use the precision matrix from the ICAR model $\mathbf{Q} = (\text{diag}(\mathbf{W1}) - \mathbf{W})$. We also compare inferential and predictive performance across the three different precision matrices as in the binary case.

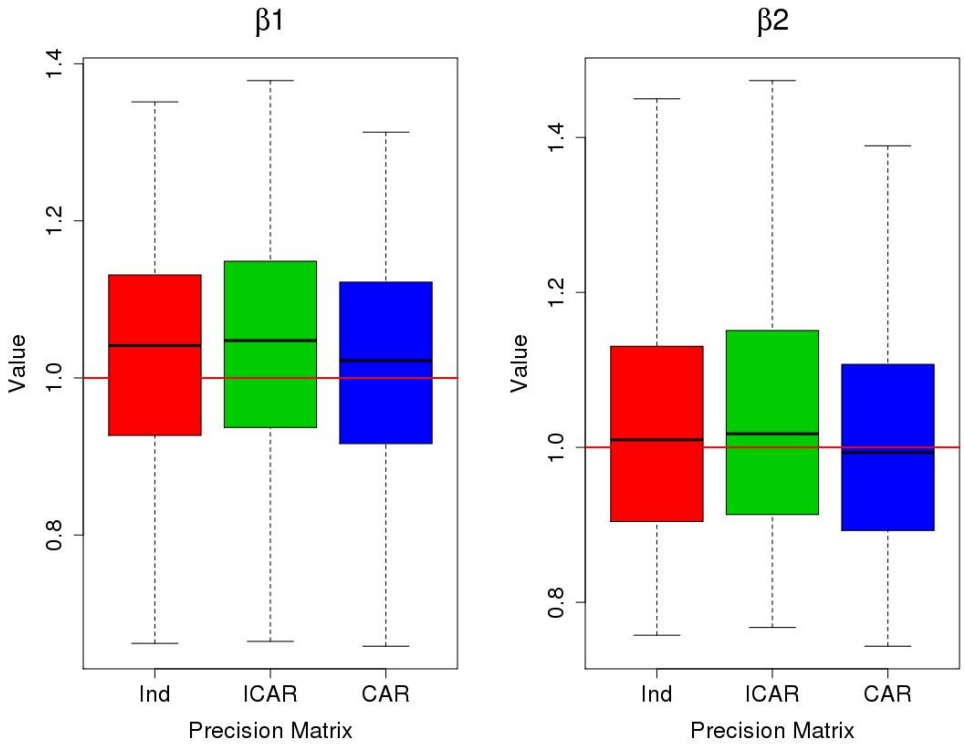

Figure B.2. Binary data simulation study: distribution of posterior mean estimates for parameters β_1 (left) and β_2 (right) for three different precision matrices - Independent (red), ICAR (green), and CAR with $\phi = 0.5$ (blue). The suitable rank p of the Moran's operator $\mathbf{M}$ chosen using the automated heuristic. Distributions are similar across precision matrices.

Results indicate that the choice of rank (for the Moran's operator) is a key driver for accurate parameter estimation and prediction as noted in Table B.2. As in the binary case, the choice of precision matrices does not influence inference or prediction as shown in Table B.3. Coverage probabilities (Table B.4) align with the nominal coverage (95%). The PICAR approach improves mixing in the MCMC algorithm as shown by the larger effective samples per second (ESS/sec) compared to the gold standard approach. For model parameters β_1 and β_2, PICAR yields an ESS/sec of 6.4 and 7.2 respectively and the gold standard returns an ESS/sec 0.09 and 0.09 respectively. For the random effects $\mathbf{W}$, the average ESS/sec is 1.8 for the PICAR approach and 0.018 for the gold standard, an improvement by a factor of roughly 101.

Rank	β_1 (95% CI)	β_2 (95% CI)	CVMPSE	Time (min)
10	1.09 (0.99,1.19)	1.01 (0.92,1.11)	1.96	8.84
50	1.05 (0.95,1.15)	1.02 (0.92,1.12)	1.74	9.87
62	**1.04 (0.94,1.14)**	**0.99 (0.89,1.09)**	**1.57**	**10.65**
75	1.03 (0.93,1.14)	0.99 (0.89,1.09)	1.66	10.39
100	1.05 (0.95,1.16)	0.98 (0.88,1.09)	1.71	11.07
200	1.08 (0.97,1.19)	0.98 (0.87,1.1)	1.81	13.49
Gold Standard	1.07 (0.97,1.17)	1.01 (0.91,1.12)	1.66	3803.84

Table B.2. Simulated example with count spatial observations. Parameter estimation, prediction, and model fitting time results across Moran's basis ranks. Bold font denotes the rank chosen by the automated heuristic.

Precision Matrix	β_1 (95% CI)	β_2 (95% CI)	CVMPSE	Time (min)
Ind	1.04 (0.94,1.14)	0.99 (0.89,1.09)	1.56	10.92
ICAR	1.04 (0.94,1.14)	0.99 (0.89,1.09)	1.57	10.65
CAR	1.04 (0.94,1.15)	0.99 (0.89,1.09)	1.57	10.17
Gold Standard	1.07 (0.97,1.17)	1.01 (0.91,1.12)	1.66	3803.84

Table B.3. Simulated example with count spatial observations. Parameter estimation, prediction, and model fitting time results across precision matrices.

	β_1	β_2
Independent	0.95	0.97
ICAR	0.95	0.96
CAR	0.95	0.95

Table B.4. Poisson data simulation study: Coverage probabilities for 100 simulated samples. Columns correspond to the regression coefficients. Rows correspond to the type of precision matrix.

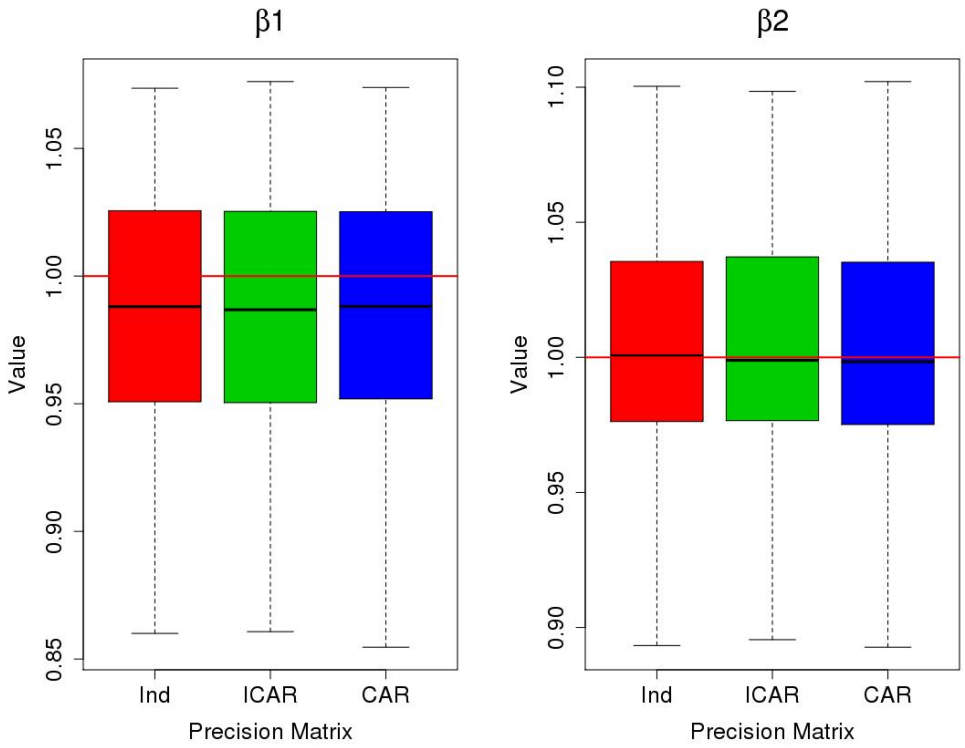

Figure B.3. Poisson data simulation study: distribution of posterior mean estimates for parameters β_1 (left) and β_2 (right) for three different precision matrices - Independent (red), ICAR (green), and CAR with $\phi = 0.5$ (blue). The suitable rank p of the Moran's operator $\mathbf{M}$ chosen using the automated heuristic. Distributions are similar across precision matrices.

	1	2	3	4	5	6
Rank	β_1 (95% CI)	β_2 (95% CI)	α_1 (95% CI)	α_2 (95% CI)	Mismatch %	Time (min)
10	0.88 (0.64,1.11)	1.01 (0.77,1.24)	-0.03 (-0.17,0.11)	-0.02 (-0.2,0.15)	0.43	12.77
23	**0.9 (0.67,1.15)**	**1.05 (0.81,1.28)**	**0 (-0.14,0.14)**	**0 (-0.17,0.17)**	**0.42**	**12.64**
50	0.94 (0.69,1.18)	1.12 (0.87,1.36)	0.05 (-0.09,0.18)	0.05 (-0.13,0.22)	0.41	13.64
75	0.98 (0.73,1.23)	1.17 (0.92,1.43)	0.07 (-0.07,0.21)	0.07 (-0.1,0.25)	0.41	14.63
100	1.01 (0.77,1.27)	1.22 (0.96,1.48)	0.09 (-0.05,0.22)	0.09 (-0.09,0.26)	0.4	16.7
200	1.12 (0.85,1.4)	1.24 (0.97,1.52)	0.15 (0.01,0.29)	0.13 (-0.05,0.3)	0.41	18.92
300	1.23 (0.94,1.52)	1.35 (1.06,1.64)	0.22 (0.08,0.35)	0.19 (0.02,0.36)	0.41	20.94
Gold Standard	0.9 (0.65,1.12)	1.07 (0.82,1.31)	0.02 (-0.12,0.16)	0.02 (-0.17,0.18)	0.42	7558.88

Table B.5. Simulated example with ordered categorical spatial observations. Parameter estimation, prediction, and model fitting time results across Moran's basis ranks. Bold font denotes the rank chosen by the automated heuristic.

	1	2	3	4	5	6
Rank	β_1 (95% CI)	β_2 (95% CI)	α_1 (95% CI)	α_2 (95% CI)	Mismatch %	Time (min)
Ind	0.9 (0.66,1.14)	1.04 (0.8,1.28)	0.01 (-0.13,0.14)	0 (-0.17,0.17)	0.42	11.70
ICAR	0.9 (0.67,1.15)	1.05 (0.81,1.28)	0 (-0.14,0.14)	0 (-0.17,0.17)	0.42	12.63
CAR	0.89 (0.65,1.12)	1.03 (0.79,1.26)	0.01 (-0.13,0.15)	0 (-0.17,0.18)	0.43	12.87
Gold Standard	0.9 (0.65,1.12)	1.07 (0.82,1.31)	0.02 (-0.12,0.16)	0.02 (-0.17,0.18)	0.42	7558.88

Table B.6. Simulated example with ordered categorical spatial observations. Parameter estimation, prediction, and model fitting time results across precision matrices.

	β_1	β_2	α_1	α_2
Independent	0.92	0.94	0.93	0.90
ICAR	0.91	0.92	0.93	0.88
CAR	0.93	0.96	0.93	0.94

Table B.7. Ordered categorical data simulation study: Coverage probabilities for 100 simulated samples. Columns correspond to the regression coefficients. Rows correspond to the type of precision matrix.

Covariate	Estimate	95% CI
Age	0.0008	(-0.0041,0.0034)
Basal Area	-0.0045	(-0.007,-0.0026)
Height	0.0203	(0.0157,0.0236)
Volume	-0.0026	(-0.0034,-0.0017)
tau	0.0040	(0.0021,0.0094)

Table B.8. Inference results for the mistletoe data. Rows correspond to the predictor variables and columns include the parameter estimates and 95% credible intervals

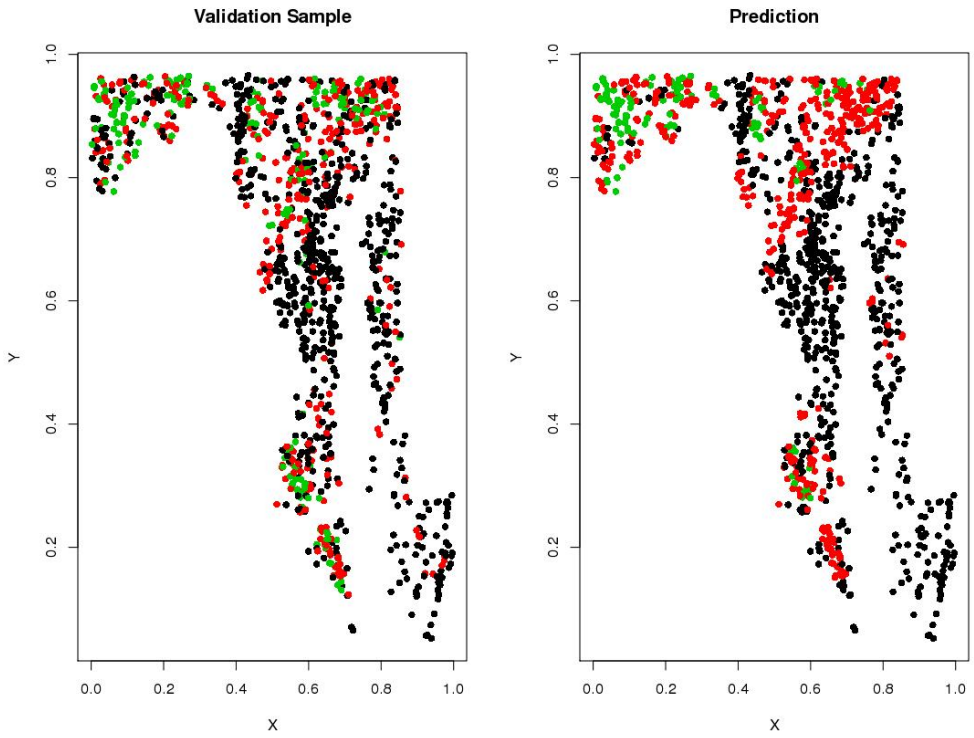

Figure B.4. The left panel shows the BIBI index at the prediction locations and the right panel shows the predicted BIBI index. Black, red, and green points indicate low, medium and high levels of BIBI respectively.

Bibliography

Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D., Stuart, A., et al. (2017). Importance sampling: Intrinsic dimension and computational cost. *Statistical Science*, 32(3):405–431.

Agarwal, D. K., Gelfand, A. E., and Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics*, 9(4):341–355.

Agresti, A. (2010). *Analysis of ordinal categorical data*, volume 656. John Wiley & Sons.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679.

Albert, J. H. and Chib, S. (1997). Bayesian methods for cumulative, sequential, and two-step ordinal data regression models. Technical report, Bowling Green State University.

Alder, J. R., Hostetler, S. W., Pollard, D., and Schmittner, A. (2011). Evaluation of a present-day climate simulation with a new coupled atmosphere-ocean model GENMOM. *Geoscientific Model Development*, 4(1):69–83.

Applegate, P. J., Kirchner, N., Stone, E. J., Keller, K., and Greve, R. (2012). An assessment of key model parametric uncertainties in projections of Greenland ice sheet behavior. *The Cryosphere*, 6(3):589–606.

Arcuti, S., Pollice, A., Ribecco, N., and D'Onghia, G. (2016). Bayesian spatiotemporal analysis of zero-inflated biological population density data by a delta-normal spatiotemporal additive model: Bayesian analysis of zero-inflated biological data. *Biometrical Journal*, 58(2):372–386.

Bakker, A. M., Applegate, P. J., and Keller, K. (2016). A simple, physically motivated model of sea-level contributions from the Greenland ice sheet in response to temperature changes. *Environmental Modelling & Software*, 83:27–35.

Ballard, G., Siefert, C., and Hu, J. (2016). Reducing communication costs for sparse matrix multiplication within algebraic multigrid. *SIAM Journal on Scientific Computing*, 38(3):C203–C231.

Banerjee, A., Dunson, D. B., and Tokdar, S. T. (2013). Efficient Gaussian process regression for large datasets. *Biometrika*, 100(1):75–89.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.

Bastos, L. S. and O'Hagan, A. (2009). Diagnostics for Gaussian process emulators. *Technometrics*, 51(4):425–438.

Bates, D. and Maechler, M. (2019). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-17.

Bayarri, M., Berger, J., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R., Paulo, R., Sacks, J., Walsh, D., et al. (2007). Computer model validation with functional output. *The Annals of Statistics*, 35(5):1874–1906.

Berliner, L. M. (1996). Hierarchical Bayesian time series models. In *Maximum entropy and Bayesian methods*, pages 15–22. Springer.

Bernadinelli, L., Pascutto, C., Best, N., and Gilks, W. (1997). Disease mapping with errors in covariates. *Statistics in Medicine*, 16(7):741–752.

Bernton, E., Yang, S., Chen, Y., Shephard, N., and Liu, J. S. (2015). Locally weighted markov chain monte carlo. *arXiv preprint arXiv:1506.08852*.

Berveiller, M., Sudret, B., and Lemaire, M. (2006). Stochastic finite element: a non intrusive approach by regression. *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique*, 15(1-3):81–92.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1):1–20.

Beskos, A., Jasra, A., Kantas, N., Thiery, A., et al. (2016). On the convergence of adaptive sequential Monte Carlo methods. *The Annals of Applied Probability*, 26(2):1111–1146.

Bhat, K. S., Haran, M., Goes, M., and Chen, M. (2010). Computer model calibration with multivariate spatial output: A case study. *Frontiers of Statistical Decision Making and Bayesian Analysis*, pages 168–184.

Bhattacharyya, A. (1946). On a measure of divergence between two multinomial populations. *Sankhyā: The Indian Journal of Statistics*, pages 401–406.

Bijleveld, A. I., van Gils, J. A., van der Meer, J., Dekinga, A., Kraan, C., van der Veer, H. W., and Piersma, T. (2012). Designing a benthic monitoring programme with multiple conflicting objectives. *Methods in Ecology and Evolution*, 3(3):526–536.

Bony, S. and Emanuel, K. A. (2001). A parameterization of the cloudiness associated with cumulus convection; evaluation using TOGA COARE data. *Journal of the Atmospheric Sciences*, 58(21):3158–3183.

Bopp, G. P., Shaby, B. A., Forest, C. E., and Mejía, A. (2020). Projecting flood-inducing precipitation with a bayesian analogue model. *Journal of Agricultural, Biological and Environmental Statistics*, 25(2):229–249.

Bradley, J. R., Cressie, N., Shi, T., et al. (2016). A comparison of spatial predictors when datasets could be very large. *Statistics Surveys*, 10:100–131.

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2019). Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family. *Journal of the American Statistical Association*, 0(ja):1–29.

Brenner, S. and Scott, R. (2007). *The mathematical theory of finite element methods*, volume 15. Springer Science & Business Media.

Brynjarsdóttir, J. and Berliner, L. M. (2014). Dimension-reduced modeling of spatio-temporal processes. *Journal of the American Statistical Association*, 109(508):1647–1659.

Brynjarsdottir, J. and O'Hagan, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30(11):114007.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1).

Chang, W., Haran, M., Applegate, P., and Pollard, D. (2016a). Calibrating an ice sheet model using high-dimensional binary spatial data. *Journal of the American Statistical Association*, 111(513):57–72.

Chang, W., Haran, M., Applegate, P., Pollard, D., et al. (2016b). Improving ice sheet model calibration using paleoclimate and modern data. *The Annals of Applied Statistics*, 10(4):2274–2302.

Chang, W., Haran, M., Olson, R., Keller, K., et al. (2014). Fast dimension-reduced climate model calibration and the effect of data aggregation. *The Annals of Applied Statistics*, 8(2):649–673.

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, 89(3):539–552.

Christensen, O. F., Roberts, G. O., and Sköld, M. (2006). Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15(1):1–17.

Ciccazzo, A., Di Pillo, G., and Latorre, V. (2014). Support vector machines for surrogate modeling of electronic circuits. *Neural Computing and Applications*, 24(1):69–76.

Collins, M. (2007). Ensembles and probabilities: A new era in the prediction of climate change. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1857):1957–1970.

Compton, T. J., Holthuijsen, S., Koolhaas, A., Dekinga, A., ten Horn, J., Smith, J., Galama, Y., Brugge, M., van der Wal, D., van der Meer, J., et al. (2013). Distinctly variable mudscapes: distribution gradients of intertidal macrofauna across the dutch wadden sea. *Journal of Sea Research*, 82:103–116.

Computational and Information Systems Laboratory (2017). Cheyenne: HPE/SGI ICE XA System (University Community Computing). Boulder, CO: National Center for Atmospheric Research. doi:10.5065/D6RX99HX.

Cook, C., Hill, D., van de Flierdt, T., Williams, T., Hemming, S., Dolan, A., Pierce, E., Escutia, C., Harwood, D., Cortese, G., et al. (2014). Sea surface temperature control on the distribution of far-traveled Southern Ocean ice-rafted detritus during the Pliocene. *Paleoceanography*, 29(6):533–548.

Cook, C. P., van de Flierdt, T., Williams, T., Hemming, S. R., Iwai, M., Kobayashi, M., Jimenez-Espejo, F. J., Escutia, C., González, J. J., Khim, B.-K., McKay, R. M., Passchier, S., Bohaty, S. M., Riesselman, C. R., Tauxe, L., Sugisaki, S., Galindo, A. L., Patterson, M. O., Sangiorgi, F., Pierce, E. L., Brinkhuis, H., Klaus, A., Fehr, A., Bendle, J. A. P., Bijl, P. K., Carr, S. A., Dunbar, R. B., Flores, J. A., Hayden, T. G., Katsuki, K., Kong, G. S., Nakai, M., Olney, M. P., Pekar, S. F., Pross, J., Röhl, U., Sakai, T., Shrivastava, P. K., Stickley, C. E., Tuo, S., Welsh, K., and Yamane, M. (2013). Dynamic behaviour of the East Antarctic ice sheet during Pliocene warmth. *Nature Geoscience*, 6(9):765–769.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.

Cressie, N. and Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Crisan, D. and Doucet, A. (2000). Convergence of sequential Monte Carlo methods. *Signal Processing Group, Department of Engineering, University of Cambridge, Technical Report CUED/F-INFENG/TR381*, 1.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416):953–963.

Dal Gesso, S. and Neggers, R. (2018). Can we use single-column models for understanding the boundary layer cloud-climate feedback? *Journal of Advances in Modeling Earth Systems*, 10(2):245–261.

Dancik, G. M. and Dorman, K. S. (2008). mlegp: statistical analysis for computer models of biological systems using r. *Bioinformatics*, 24(17):1967.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812.

De Boer, B., Stocchi, P., and Van De Wal, R. (2014). A fully coupled 3-D ice-sheet-sea-level model: Algorithm and applications. *Geoscientific Model Development*, 7(5):2141–2156.

De Oliveira, V. (2000). Bayesian prediction of clipped Gaussian random fields. *Computational Statistics & Data Analysis*, 34(3):299–314.

de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., and Bodik, R. (2017). Programming with models: writing statistical algorithms for general model structures with nimble. *Journal of Computational and Graphical Statistics*, 26:403–413.

DeConto, R. M. and Pollard, D. (2016). Contribution of Antarctica to past and future sea-level rise. *Nature*, 531(7596):591.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B*, 68(3):411–436.

Deschamps, P., Durand, N., Bard, E., Hamelin, B., Camoin, G., Thomas, A. L., Henderson, G. M., Okuno, J., and Yokoyama, Y. (2012). Ice-sheet collapse and sea-level rise at the Bølling warming 14,600 years ago. *Nature*, 483(7391):559.

Diaz, D. and Keller, K. (2016). A potential disintegration of the West Antarctic ice sheet: Implications for economic analyses of climate policy. *American Economic Review*, 106(5):607–11.

Diggle, P. J., Tawn, J. A., and Moyeed, R. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.

Dolan, A. M., De Boer, B., Bernales, J., Hill, D. J., and Haywood, A. M. (2018). High climate model dependency of Pliocene Antarctic ice-sheet predictions. *Nature Communications*, 9(1):2799.

Dolan, A. M., Haywood, A. M., Hill, D. J., Dowsett, H. J., Hunter, S. J., Lunt, D. J., and Pickering, S. J. (2011). Sensitivity of Pliocene ice sheets to orbital forcing. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 309(1-2):98–110.

Doucet, A., De Freitas, N., and Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo Methods in Practice*, pages 3–14. Springer.

Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208.

Dowsett, H. J. and Cronin, T. M. (1990). High eustatic sea level during the middle Pliocene: Evidence from the southeastern US Atlantic Coastal Plain. *Geology*, 18(5):435–438.

Dreassi, E., Petrucci, A., and Rocco, E. (2014). Small area estimation for semicontinuous skewed spatial data: An application to the grape wine production in tuscany: Small area estimation for semicontinuous skewed spatial data. *Biometrical Journal*, 56(1):141–156.

Drovandi, C. C., Pettitt, A. N., and Faddy, M. J. (2011). Approximate Bayesian computation using indirect inference. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(3):317–337.

Dutton, A., Carlson, A., Long, A., Milne, G., Clark, P., DeConto, R., Horton, B., Rahmstorf, S., and Raymo, M. (2015). Sea-level rise due to polar ice-sheet mass loss during past warm periods. *Science*, 349(6244):aaa4019.

Eason, J. and Cremaschi, S. (2014). Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Computers & Chemical Engineering*, 68:220–232.

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(7742):58.

Eldred, M., Webster, C., and Constantine, P. (2008). Evaluation of non-intrusive approaches for wiener-askey generalized polynomial chaos. In *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 16th AIAA/ASME/AHS Adaptive Structures Conference, 10th AIAA Non-Deterministic Approaches Conference, 9th AIAA Gossamer Spacecraft Forum, 4th AIAA Multidisciplinary Design Optimization Specialists Conference*, page 1892.

Fan, W., Yu, W., Xu, J., Zhou, J., Luo, X., Yin, Q., Lu, P., Cao, Y., and Xu, R. (2018). Parallelizing sequential graph computations. *ACM Transactions on Database Systems (TODS)*, 43(4):18.

Fernandes, M. V., Schmidt, A. M., and Migon, H. S. (2009). Modelling zero-inflated spatio-temporal processes. *Statistical Modelling: An International Journal*, 9(1):3–25.

Fitzgerald, P., Bamber, J., Ridley, J., and Rougier, J. (2012). Exploration of parametric uncertainty in a surface mass balance model applied to the Greenland ice sheet. *Journal of Geophysical Research: Earth Surface*, 117(F1).

Fretwell, P., Pritchard, H., Vaughan, D., Bamber, J., Barrand, N., Bell, R., Bianchi, C., Bingham, R., Blankenship, D., Casassa, G., et al. (2012). Bedmap2: Improved ice bed, surface and thickness datasets for Antarctica. *The Cryosphere Discussions*, 6:4305–4361.

Fruhwirth-Schnatter, S. and Pyne, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics*, 11(2):317–336.

Fuentes, M. (2001). A high frequency kriging approach for non-stationary environmental processes. *Environmetrics: The official journal of the International Environmetrics Society*, 12(5):469–483.

Fuller, R. W., Wong, T. E., and Keller, K. (2017). Probabilistic inversion of expert assessments to inform projections about Antarctic ice sheet responses. *PloS One*, 12(12):e0190115.

Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.

Garner, G. G. and Keller, K. (2018). Using direct policy search to identify robust strategies in adapting to uncertain sea-level rise and storm surge. *Environmental Modelling & Software*, 107:96–104.

Geils, B. and Hawksworth, F. (2002). Damage, effects, and importance of dwarf mistletoes. *In: Geils, Brian W.; Cibrián Tovar, Jose; Moody, Benjamin, tech. coords. Mistletoes of North American Conifers. Gen. Tech. Rep. RMRS-GTR-98. Ogden, UT: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 57-65*, 98.

Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396.

George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Gettelman, A., Truesdale, J. E., Bacmeister, J. T., Caldwell, P. M., Neale, R. B., Bogenschutz, P. A., and Simpson, I. R. (2019). The single column atmosphere model version 6 (SCAM6): Not a scam but a tool for model evaluation and development. *Journal of Advances in Modeling Earth Systems*, 0(ja).

Ghanem, R. G. and Spanos, P. D. (1991). Spectral stochastic finite-element formulation for reliability analysis. *Journal of Engineering Mechanics*, 117(10):2351–2372.

Ghiocel, D. M. and Ghanem, R. G. (2002). Stochastic finite-element analysis of seismic soil–structure interaction. *Journal of Engineering Mechanics*, 128(1):66–77.

Gilks, W. R. and Berzuini, C. (2001). Following a moving target—Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(1):127–146.

Giraud, F., Del Moral, P., et al. (2017). Nonasymptotic analysis of adaptive and annealed Feynman–Kac particle models. *Bernoulli*, 23(1):670–709.

Golledge, N. R., Keller, E. D., Gomez, N., Naughten, K. A., Bernales, J., Trusel, L. D., and Edwards, T. L. (2019). Global environmental consequences of twenty-first-century ice-sheet melt. *Nature*, 566(7742):65.

Goodman, J. and Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in applied mathematics and computational science*, 5(1):65–80.

Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F-radar and Signal Processing*, volume 140, pages 107–113. IET.

Gorissen, D., De Tommasi, L., Crombecq, K., and Dhaene, T. (2009). Sequential modeling of a low noise amplifier with neural networks and active learning. *Neural Computing and Applications*, 18(5):485–494.

Gramacy, R., Samworth, R., and King, R. (2010). Importance tempering. *Statistics and Computing*, 20(1):1–7.

Gramacy, R. B. and Apley, D. W. (2015). Local Gaussian process approximation for large computer experiments. *Journal of Computational and Graphical Statistics*, 24(2):561–578.

Greve, R. (1997). Application of a polythermal three-dimensional ice sheet model to the Greenland ice sheet: Response to steady-state and transient climate scenarios. *Journal of Climate*, 10(5):901–918.

Griffith, D. A. (2003). Spatial filtering. In *Spatial Autocorrelation and Spatial Filtering*, pages 91–130. Springer.

Gschlößl, S. and Czado, C. (2008). Modelling count data with overdispersion and spatial effects. *Statistical Papers*, 49(3):531–552.

Guan, Y. and Haran, M. (2018). A computationally efficient projection-based approach for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 27(4):701–714.

Guan, Y. and Haran, M. (2019). Fast expectation-maximization algorithms for spatial generalized linear mixed models. *arXiv preprint arXiv:1909.05440*.

Guhaniyogi, R., Finley, A. O., Banerjee, S., and Gelfand, A. E. (2011). Adaptive Gaussian predictive process models for large spatial datasets. *Environmetrics*, 22(8):997–1007.

Hall, D. B. (2000). Zero-inflated poisson and binomial regression with random effects: a case study. *Biometrics*, 56(4):1030–1039.

Hanks, E. M., Hooten, M. B., and Baker, F. A. (2011). Reconciling multiple data sources to improve accuracy of large-scale prediction of forest disease incidence. *Ecological Applications*, 21(4):1173–1188.

Hannart, A., Ghil, M., Dufresne, J.-L., and Naveau, P. (2013). Disconcerting learning on climate sensitivity and the uncertain future of uncertainty. *Climatic Change*, 119(3-4):585–601.

Haran, M. (2011). Gaussian random field models for spatial data. *Handbook of Markov Chain Monte Carlo*, pages 449–478.

Haran, M., Hodges, J. S., and Carlin, B. P. (2003). Accelerating computation in markov random field models for spatial data via structured MCMC. *Journal of Computational and Graphical Statistics*, 12(2):249–264.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., et al. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24(3):398–425.

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, 5(2):173–190.

Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482):570–583.

Higdon, D., Swall, J., and Kern, J. (1999). Non-stationary spatial modeling. bayesian statistics 6. In *Proceedings of the Sixth Valencia International Meeting*, pages 761–768.

Higgs, M. D. and Hoeting, J. A. (2010). A clipped latent variable model for spatially correlated ordered categorical data. *Computational Statistics & Data Analysis*, 54(8):1999–2011.

Hjelle, Ø. and Dæhlen, M. (2006). *Triangulations and applications*. Springer Science & Business Media.

Hoef, J. M. V. and Jansen, J. K. (2007). Space—time zero-inflated count models of harbor seals. *Environmetrics*, 18(7):697–712.

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Hu, M.-C., Pavlicova, M., and Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: examples from an hiv-risk reduction intervention trial. *The American journal of drug and alcohol abuse*, 37(5):367–375.

Hughes, J. and Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):139–159.

Isaac, T., Petra, N., Stadler, G., and Ghattas, O. (2015a). Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet. *Journal of Computational Physics*, 296:348–368.

Isaac, T., Stadler, G., and Ghattas, O. (2015b). Solution of nonlinear stokes equations discretized by high-order finite elements on nonconforming and anisotropic meshes, with application to ice sheet dynamics. *SIAM Journal on Scientific Computing*, 37(6):B804–B833.

Ishwaran, H., Rao, J. S., et al. (2005). Spike and slab variable selection: frequentist and bayesian strategies. *The Annals of Statistics*, 33(2):730–773.

Isukapalli, S., Roy, A., and Georgopoulos, P. (1998). Stochastic response surface methods (srsms) for uncertainty propagation: application to environmental and biological systems. *Risk analysis*, 18(3):351–363.

Jackson, C. H., Jit, M., Sharples, L. D., and De Angelis, D. (2015). Calibration of complex models through Bayesian evidence synthesis: A demonstration and tutorial. *Medical Decision Making*, 35(2):148–161.

Jacobs, S. S., Jenkins, A., Giulivi, C. F., and Dutrieux, P. (2011). Stronger ocean circulation and increased melting under Pine Island Glacier ice shelf. *Nature Geoscience*, 4:519.

Jandarov, R., Haran, M., Bjørnstad, O., and Grenfell, B. (2014). Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63(3):423–444.

Jasra, A., Stephens, D. A., Doucet, A., and Tsagaris, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scandinavian Journal of Statistics*, 38(1):1–22.

Jeremiah, E., Sisson, S., Marshall, L., Mehrotra, R., and Sharma, A. (2011). Bayesian calibration and uncertainty analysis of hydrological models: A comparison of adaptive Metropolis and sequential Monte Carlo samplers. *Water Resources Research*, 47(7).

Johnson, D. R., Fischbach, J. R., and Ortiz, D. S. (2013). Estimating surge-based flood risk with the coastal louisiana risk assessment model. *Journal of Coastal Research*, 67(sp1):109–126.

Johnson, V. E. and Albert, J. H. (2006). *Ordinal data modeling*. Springer Science & Business Media.

Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). Fixed-width output analysis for markov chain monte carlo. *Journal of the American Statistical Association*, 101(476):1537–1547.

Joyce, P. and Marjoram, P. (2008). Approximately sufficient statistics and Bayesian computation. *Statistical applications in genetics and molecular biology*, 7(1).

Kalyanaraman, J., Kawajiri, Y., Lively, R. P., and Realff, M. J. (2016). Uncertainty quantification via Bayesian inference using sequential Monte Carlo methods for CO2 adsorption process. *AIChE Journal*, 62(9):3352–3368.

Kantas, N., Beskos, A., and Jasra, A. (2014). Sequential Monte Carlo methods for high-dimensional inverse problems: A case study for the Navier–Stokes equations. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):464–489.

Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., Chopin, N., et al. (2015). On particle methods for parameter estimation in state-space models. *Statistical Science*, 30(3):328–351.

Katzfuss, M. (2017). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association*, 112(517):201–214.

Keller, K. and McInerney, D. (2008). The dynamics of learning about a climate threshold. *Climate Dynamics*, 30(2-3):321–332.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B*, 63(3):425–464.

Kim, S. H., Chang, C.-C. H., Kim, K. H., Fine, M. J., and Stone, R. A. (2012). Blup (remql) estimation of a correlated random effects negative binomial hurdle model. *Health Services and Outcomes Research Methodology*, 12(4):302–319.

Kim, S. H., Edmonds, J., Lurz, J., Smith, S. J., and Wise, M. (2006). The objECTS framework for integrated assessment: Hybrid modeling of transportation. *The Energy Journal*, pages 63–91.

Kong, A. (1992). A note on importance sampling using standardized weights. *University of Chicago, Dept. of Statistics, Tech. Rep*, 348.

Kopp, R. E., Simons, F. J., Mitrovica, J. X., Maloof, A. C., and Oppenheimer, M. (2009). Probabilistic assessment of sea level during the Last Interglacial stage. *Nature*, 462(7275):863.

Lambert, D. (1992a). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14.

Lambert, D. (1992b). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1.

Larour, E., Seroussi, H., Morlighem, M., and Rignot, E. (2012). Continental scale, high order, high spatial resolution, ice sheet modeling using the Ice Sheet System Model (ISSM). *Journal of Geophysical Research: Earth Surface*, 117(F1).

Le Bars, D., Drijfhout, S., and de Vries, H. (2017). A high-end sea level rise probabilistic projection including rapid Antarctic ice sheet mass loss. *Environmental Research Letters*, 12(4):044013.

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 System Science Data*, 2(2):247–260.

Le Cozannet, G., Manceau, J.-C., and Rohmer, J. (2017). Bounding probabilistic sea-level projections within the framework of the possibility theory. *Environmental Research Letters*, 12(1):014012.

Le Maıtre, O. P., Reagan, M. T., Najm, H. N., Ghanem, R. G., and Knio, O. M. (2002). A stochastic projection method for fluid flow: Ii. random process. *Journal of computational Physics*, 181(1):9–44.

Lee, B. S. and Haran, M. (2019). Picar: An efficient extendable approach for fitting hierarchical spatial models. *arXiv preprint arXiv:1912.02382*.

Lee, B. S., Haran, M., Fuller, R., Pollard, D., and Keller, K. (2019). Supplement to "Supplement: A Fast Particle-based Approach for Calibrating a 3-D Model of the Antarctic Ice Sheet".

Lee, B. S., Haran, M., Fuller, R. W., Pollard, D., and Keller, K. (2020+). A fast particle-based approach for calibrating a 3-D model of the Antarctic ice sheet. to appear in the Annals of Applied Statistics.

Lee, C.-E. and Kim, S. (2017). Applicability of zero-inflated models to fit the torrential rainfall count data with extra zeros in south korea. *Water*, 9(2):123.

Lee, Y., Alam, M. M., Noh, M., Rönnegård, L., and Skarin, A. (2016). Spatial modeling of data with excessive zeros applied to reindeer pellet-group counts. *Ecology and evolution*, 6(19):7047–7056.

Lehoucq, R. B., Sorensen, D. C., and Yang, C. (1998). *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*, volume 6. Siam.

Lele, S. R., Dennis, B., and Lutscher, F. (2007). Data cloning: Easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, 10(7):551–563.

Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R. A., Mishonov, A. V., Reagan, J., Seidov, D., Yarosh, E. S., et al. (2012). World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophysical Research Letters*, 39(10).

Li, T., Sun, S., Sattar, T. P., and Corchado, J. M. (2014). Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches. *Expert Systems with Applications*, 41(8):3944–3954.

Liang, F. and Wong, W. H. (2001). Real-Parameter Evolutionary Monte Carlo with applications to Bayesian mixture models. *Journal of the American Statistical Association*, 96(454):653–666.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, 99(D7):14415–14428.

Lindgren, F., Rue, H., et al. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19):1–25.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Liu, F., Bayarri, M. J., and Berger, J. O. (2009a). Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1):119–150. Zbl: 1330.65033.

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice*, pages 197–223. Springer.

Liu, J. S. and Chen, R. (1998). Sequential monte carlo methods for dynamic systems. *Journal of the American statistical association*, 93(443):1032–1044.

Liu, J. S., Liang, F., and Wong, W. H. (2000). The multiple-try method and local optimization in Metropolis sampling. *Journal of the American Statistical Association*, 95(449):121–134.

Liu, L., Strawderman, R. L., Johnson, B. A., and O'Quigley, J. M. (2016). Analyzing repeated measures semi-continuous data, with application to an alcohol dependence study. *Statistical Methods in Medical Research*, 25(1):133–152.

Liu, X. and Guillas, S. (2017). Dimension reduction for Gaussian process emulation: An application to the influence of bathymetry on tsunami heights. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):787–812.

Liu, Z., Otto-Bliesner, B., He, F., Brady, E., Tomas, R., Clark, P., Carlson, A., Lynch-Stieglitz, J., Curry, W., Brook, E., et al. (2009b). Transient simulation of last deglaciation with a new mechanism for Bølling-Allerød warming. *Science*, 325(5938):310–314.

Llopis, F. P., Kantas, N., Beskos, A., and Jasra, A. (2018). Particle filtering for ftochastic Navier–Stokes signal observed with linear additive noise. *SIAM Journal on Scientific Computing*, 40(3):A1544–A1565.

Lorenz, E. N. (1972). Predictability: Does the flap of a butterfly's wings in Brazil set off a tornado in Texas? Presented at the 139th Meeting of the AAAS. http://eaps4.mit.edu/research/Lorenz/Butterfly_1972.pdf. Accessed: 2019-08-04.

Lyashevska, O., Brus, D. J., and van der Meer, J. (2016). Mapping species abundance by a spatial zero-inflated poisson model: a case study in the wadden sea, the netherlands. *Ecology and evolution*, 6(2):532–543.

Maniyar, D., Cornford, D., and Boukouvalas, A. (2007). Dimensionality reduction in the emulator setting. Technical report, Neural Computing Research Group, University of Aston.

Martino, L. (2018). A review of multiple try MCMC algorithms for signal processing. *Digital Signal Processing*, 75:134–152.

Martino, L., Elvira, V., and Louzada, F. (2017). Effective sample size for importance sampling based on discrepancy measures. *Signal Processing*, 131:386–401.

Maryland's Mapping and GIS Data Portal (2018). Maryland Stream Health - Stream Wader Sites volunteer collected. Data retrieved from MD iMAP, `https://geodata.md.gov/imap/rest/services/Hydrology/MD_StreamHealth/FeatureServer/0`.

McGranahan, G., Balk, D., and Anderson, B. (2007). The rising tide: Assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and Urbanization*, 19(1):17–37.

McKay, M. D., Beckman, R. J., and Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M., Lamarque, J.-F., Matsumoto, K., Montzka, S., Raper, S., Riahi, K., et al. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, 109(1-2):213.

Min, Y. and Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical modelling*, 5(1):1–19.

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the american statistical association*, 83(404):1023–1032.

Monier, E., Scott, J., Sokolov, A., Forest, C., and Schlosser, C. (2013). An integrated assessment modelling framework for uncertainty studies in global and regional climate change: the MIT IGSM-CAM (version 1.0). *Geoscientific Model Development Discussions*, 6(1).

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.

Morzfeld, M., Day, M. S., Grout, R. W., Heng Pau, G. S., Finsterle, S. A., and Bell, J. B. (2018). Iterative importance sampling algorithms for parameter estimation. *SIAM Journal on Scientific Computing*, 40(2):B329–B352.

Morzfeld, M., Tu, X., Wilkening, J., and Chorin, A. (2015). Parameter estimation by implicit sampling. *Communications in Applied Mathematics and Computational Science*, 10(2):205–225.

Mu, J., Wang, G., and Wang, L. (2018). Estimation and inference in spatially varying coefficient models. *Environmetrics*, 29(1):e2485.

Muff, S., Riebler, A., Held, L., Rue, H., and Saner, P. (2015). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(2):231–252.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3):341–365.

Murray, L. M., Lee, A., and Jacob, P. E. (2016). Parallel resampling in the particle filter. *Journal of Computational and Graphical Statistics*, 25(3):789–805.

Murray, N. L., Holmes, H. A., Liu, Y., and Chang, H. H. (2019). A bayesian ensemble approach to combine pm2. 5 estimates from statistical models using satellite imagery and numerical model simulation. *Environmental research*, 178:108601.

Murray, T., Selmes, N., James, T. D., Edwards, S., Martin, I., O'Farrell, T., Aspey, R., Rutt, I., Nettles, M., and Baugé, T. (2015). Dynamics of glacier calving at the ungrounded margin of Helheim Glacier, southeast Greenland. *Journal of Geophysical Research: Earth Surface*, 120(6):964–982.

Mwalili, S. M., Lesaffre, E., and Declerck, D. (2008). The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. *Statistical methods in medical research*, 17(2):123–139.

Naish, T., Powell, R., Levy, R., Wilson, G., Scherer, R., Talarico, F., Krissek, L., Niessen, F., Pompilio, M., Wilson, T., et al. (2009). Obliquity-paced Pliocene West Antarctic ice sheet oscillations. *Nature*, 458(7236):322.

Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing*, 11(2):125–139.

Neal, R. M. (2011). Mcmc using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2.

Neelon, B., Chang, H. H., Ling, Q., and Hastings, N. S. (2016a). Spatiotemporal hurdle models for zero-inflated count data: Exploring trends in emergency department visits. *Statistical Methods in Medical Research*, 25(6):2558–2576.

Neelon, B. et al. (2018). Bayesian zero-inflated negative binomial regression based on pólya-gamma mixtures. *Bayesian Analysis*.

Neelon, B. et al. (2019). Bayesian zero-inflated negative binomial regression based on pólya-gamma mixtures. *Bayesian Analysis*, 14(3):849–875.

Neelon, B., Ghosh, P., and Loebs, P. F. (2013). A spatial poisson hurdle model for exploring geographic variation in emergency department visits: Spatial hurdle model for exploring geographic variation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2):389–413.

Neelon, B., O'Malley, A. J., and Normand, S.-L. T. (2011). A bayesian two-part latent class model for longitudinal medical expenditure data: Assessing the impact of mental health and substance abuse parity. *Biometrics*, 67(1):280–289.

Neelon, B., O'Malley, A. J., and Smith, V. A. (2016b). Modeling zero-modified count and semicontinuous data in health services research part 1: background and overview: Modeling zero-modified count and semicontinuous data in health services research part 1: background and overview. *Statistics in Medicine*, 35(27):5070–5093.

Neelon, B., Zhu, L., and Neelon, S. E. B. (2015). Bayesian two-part spatial models for semicontinuous data with application to emergency department expenditures. *Biostatistics*, 16(3):465–479.

Nguyen, T. L. T. (2014). *Sequential Monte-Carlo sampler for Bayesian inference in complex systems*. PhD thesis, Lille 1.

Nicholls, R. J., Tol, R. S., and Vafeidis, A. T. (2008). Global estimates of the impact of a collapse of the West Antarctic ice sheet: An application of FUND. *Climatic Change*, 91(1-2):171.

Nick, F. M., Veen, C. J. V. D., Vieli, A., and Benn, D. I. (2010). A physically based calving model applied to marine outlet glaciers and implications for the glacier dynamics. *Journal of Glaciology*, 56(199):781–794.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., and Sain, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24(2):579–599.

Nychka, D., Hammerling, D., Krock, M., and Wiens, A. (2018). Modeling and emulation of nonstationary gaussian fields. *Spatial statistics*, 28:21–38.

Nychka, D., Wikle, C., and Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, 2(4):315–331.

Oliver, D. S. (2003). Gaussian cosimulation: Modelling of the cross-covariance. *Mathematical Geology*, 35(6):681–698.

Olsen, M. K. and Schafer, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association*, 96(454):730–745.

O'Neill, B. C., Crutzen, P., Grübler, A., Duong, M. H., Keller, K., Kolstad, C., Koomey, J., Lange, A., Obersteiner, M., Oppenheimer, M., Pepper, W., Sanderson, W., Schlesinger, M., Treich, N., Ulph, A., Webster, M., and Wilson, C. (2006). Learning and climate change. *Climate Policy*, 6(5):585–589.

Oppenheimer, M. and Alley, R. B. (2016). How high will the seas rise? *Science*, 354(6318):1375–1377.

Owen, N. (2017). *A comparison of polynomial chaos and Gaussian process emulation for uncertainty quantification in computer experiments*. PhD thesis, University of Exeter.

O'Hagan, A. (2013). Polynomial chaos: A tutorial and critique from a statistician's perspective. *SIAM/ASA J. Uncertainty Quantification*, 20:1–20.

Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics: The official journal of the International Environmetrics Society*, 17(5):483–506.

Pal, J. S., Giorgi, F., Bi, X., Elguindi, N., Solmon, F., Gao, X., Rauscher, S. A., Francisco, R., Zakey, A., Winter, J., Ashfaq, M., Syed, F. S., Bell, J. L., Diffenbaugh, N. S., Karmacharya, J., Konaré, A., Martinez, D., da Rocha, R. P., Sloan, L. C., and Steiner, A. L. (2007). Regional climate modeling for the developing world: The ICTP RegCM3 and RegCNET. *Bulletin of the American Meteorological Society*, 88(9):1395–1410.

Papaioannou, I., Papadimitriou, C., and Straub, D. (2016). Sequential importance sampling for structural reliability analysis. *Structural Safety*, 62:66–75.

Park, J. and Haran, M. (2019). Reduced-dimensional monte carlo maximum likelihood for latent gaussian random field models. *arXiv preprint arXiv:1910.09711*.

Petra, N., Martin, J., Stadler, G., and Ghattas, O. (2014). A computational framework for infinite-dimensional Bayesian inverse problems, Part II: Stochastic Newton MCMC with application to ice sheet flow inverse problems. *SIAM Journal on Scientific Computing*, 36(4):A1525–A1555.

Pollard, D., Chang, W., Haran, M., Applegate, P., and DeConto, R. (2016). Large ensemble modeling of the last deglacial retreat of the West Antarctic ice sheet: Comparison of simple and advanced statistical techniques. *Geoscientific Model Development*, 9(5):1697–1723.

Pollard, D. and DeConto, R. (2012a). Description of a hybrid ice sheet-shelf model, and application to Antarctica. *Geoscientific Model Development*, 5(5):1273.

Pollard, D. and DeConto, R. M. (2009). Modelling West Antarctic ice sheet growth and collapse through the past five million years. *Nature*, 458(7236):329.

Pollard, D. and DeConto, R. M. (2012b). A simple inverse method for the distribution of basal sliding coefficients under ice sheets, applied to Antarctica. *The Cryosphere*, 6(5):953–971.

Pollard, D., DeConto, R. M., and Alley, R. B. (2015). Potential Antarctic Ice Sheet retreat driven by hydrofracturing and ice cliff failure. *Earth and Planetary Science Letters*, 412:112–121.

Pollard, D., Gomez, N., and Deconto, R. M. (2017). Variations of the Antarctic ice sheet in a coupled ice sheet-earth-sea level model: Sensitivity to viscoelastic earth properties. *Journal of Geophysical Research: Earth Surface*, 122(11):2124–2138.

Pruett, W. A. and Hester, R. L. (2016). The creation of surrogate models for fast estimation of complex model outcomes. *PloS one*, 11(6).

Qiu, Y. and Mei, J. (2019). *RSpectra: Solvers for Large-Scale Eigenvalue and SVD Problems*. R package version 0.15-0.

Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P. K. (2005). Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41(1):1–28.

Rasmussen, C. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA.

Rathbun, S. L. and Fei, S. (2006). A spatial zero-inflated poisson regression model for oak regeneration. *Environmental and Ecological Statistics*, 13(4):409–426.

Reagana, M. T., Najm, H. N., Ghanem, R. G., and Knio, O. M. (2003). Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection. *Combustion and Flame*, 132(3):545–555.

Recta, V., Haran, M., and Rosenberger, J. L. (2012). A two-stage model for incidence and prevalence in point-level spatial count data: Two-stage spatial model. *Environmetrics*, 23(2):162–174.

Reese, C. S., Wilson, A. G., Hamada, M., Martz, H. F., and Ryan, K. J. (2004). Integrated analysis of computer and physical experiments. *Technometrics*, 46(2):153–164.

Rignot, E., Mouginot, J., and Scheuchl, B. (2011). Ice flow of the Antarctic ice sheet. *Science*, 333(6048):1427–1430.

Roeder, K., Lynch, K. G., and Nagin, D. S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association*, 94(447):766–776.

Rovere, A., Raymo, M. E., Mitrovica, J., Hearty, P. J., O'Leary, M., and Inglis, J. (2014). The Mid-Pliocene sea-level conundrum: Glacial isostasy, eustasy and dynamic topography. *Earth and Planetary Science Letters*, 387:27–33.

Royle, J. A. and Wikle, C. K. (2005). Efficient statistical mapping of avian count data. *Environmental and Ecological Statistics*, 12(2):225–243.

Ruckert, K. L., Shaffer, G., Pollard, D., Guan, Y., Wong, T. E., Forest, C. E., and Keller, K. (2017). Assessing the impact of retreat mechanisms in a simple Antarctic ice sheet model using Bayesian calibration. *PloS One*, 12(1):e0170052.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392.

Rutt, I. C., Hagdorn, M., Hulton, N., and Payne, A. (2009). The Glimmer community ice sheet model. *Journal of Geophysical Research: Earth Surface*, 114(F2).

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical science*, pages 409–423.

Salzmann, U., Dolan, A. M., Haywood, A. M., Chan, W.-L., Voss, J., Hill, D. J., Abe-Ouchi, A., Otto-Bliesner, B., Bragg, F. J., Chandler, M. A., Contoux, C., Dowsett, H. J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D. J., Pickering, S. J., Pound, M. J., Ramstein, G., Rosenbloom, N. A., Sohl, L., Stepanek, C., Ueda, H., and Zhang, Z. (2013). Challenges in quantifying Pliocene terrestrial warming revealed by data–model discord. *Nature Climate Change*, 3(11):969–974.

Sansó, B., Forest, C. E., Zantedeschi, D., et al. (2008). Inferring climate system properties using a computer model. *Bayesian Analysis*, 3(1):1–37.

Schäfer, C. and Chopin, N. (2013). Sequential Monte Carlo on large binary sampling spaces. *Statistics and Computing*, 23(2):163–184.

Schlegel, N.-J., Seroussi, H., Schodlok, M. P., Larour, E. Y., Boening, C., Limonadi, D., Watkins, M. M., Morlighem, M., and van den Broeke, M. R. (2018). Exploration of Antarctic Ice Sheet 100-year contribution to sea level rise and associated model uncertainties using the ISSM framework. *The Cryosphere*, 12(11):3511–3534.

Schliep, E. M. and Hoeting, J. A. (2013). Multilevel latent Gaussian process model for mixed discrete and continuous multivariate response data. *Journal of agricultural, biological, and environmental statistics*, 18(4):492–513.

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(3):743–758.

Sengupta, A. and Cressie, N. (2013). Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions. *Spatial Statistics*, 4:14–44.

Sengupta, A., Cressie, N., Kahn, B. H., and Frey, R. (2016). Predictive inference for big, spatial, non-Gaussian data: MODIS cloud data and its change-of-support. *Australian & New Zealand Journal of Statistics*, 58(1):15–45.

Shaffer, G. (2014). Formulation, calibration and validation of the DAIS model, a simple Antarctic ice sheet model sensitive to variations of sea level and ocean subsurface temperature. *Geoscientific Model Development*, 7(4):1803–1818.

Shi, T. and Cressie, N. (2007). Global statistical analysis of misr aerosol data: a massive data product from nasa's terra satellite. *Environmetrics: The official journal of the International Environmetrics Society*, 18(7):665–680.

Shields, C. A. and Kiehl, J. T. (2016). Atmospheric river landfall-latitude changes in future climate simulations. *Geophysical Research Letters*, 43(16):8775–8782.

Sorooshian, S., Duan, Q., and Gupta, V. K. (1993). Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resources Research*, 29(4):1185–1194.

Sriver, R. L., Lempert, R. J., Wikman-Svahn, P., and Keller, K. (2018). Characterizing uncertain sea-level rise projections to support investment decisions. *PloS one*, 13(2):e0190641.

Stan Development Team (2019). RStan: the R interface to Stan. R package version 2.19.2.

Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143–151.

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.

Stein, M. L. (2013). Statistical properties of covariance tapers. *Journal of Computational and Graphical Statistics*, 22(4):866–885.

Steinberg, D. M. and Lin, D. K. (2006). A construction method for orthogonal Latin hypercube designs. *Biometrika*, 93(2):279–288.

Stone, E., Lunt, D., Rutt, I., Hanna, E., et al. (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(3):397–417.

Stribling, J. B., Jessup, B. K., White, J. S., Boward, D., and Hurd, M. (1998). Development of a benthic index of biotic integrity for Maryland streams. *CBWP-MANTA EA-98-3 Maryland Department of Natural Resources, Annapolis, Maryland*.

Su, L., Tom, B. D. M., and Farewell, V. T. (2009). Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics*, 10(2):374–389.

Sun, Y., Li, B., and Genton, M. G. (2012). Geostatistics for large datasets. In *Advances and Challenges in Space-Time Modelling of Natural Events*, pages 55–77. Springer.

Tibshirani, R. (1994). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288.

Urban, N. M. and Fricker, T. E. (2010). A comparison of Latin hypercube and grid ensemble designs for the multivariate emulation of an Earth system model. *Computers & Geosciences*, 36(6):746–755.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333.

Wang, X., Chen, M.-H., Kou, R., and Dey, D. (2014). Bayesian spatial-temporal modeling of ecological zero-inflated count data. *Statistica Sinica*.

Wegmann, D., Leuenberger, C., and Excoffier, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, 182(4):1207–1218.

Whiteley, N., Lee, A., Heine, K., et al. (2016). On the role of interaction in sequential monte carlo algorithms. *Bernoulli*, 22(1):494–529.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5(2):117–154.

Wikle, C. K., Zammit-Mangion, A., and Cressie, N. (2019). *Spatio-temporal Statistics with R*. CRC Press.

Willems, J. C. (1972). Dissipative dynamical systems Part I: General theory. *Archive for rational mechanics and analysis*, 45(5):321–351.

Williams, C. and Seeger, M. (2001). Using the Nyström method to speed up kernel machines. In *Advances in Neural Information Processing Systems 13*, pages 682–688. MIT Press.

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., and Yamazaki, K. (2013). History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics*, 41(7-8):1703–1729.

Wilson, P. (2015). The misuse of the vuong test for non-nested models to test for zero-inflation. *Economics Letters*, 127:51–53.

Wong, T. E., Bakker, A. M., and Keller, K. (2017). Impacts of Antarctic fast dynamics on sea-level projections and coastal flood defense. *Climatic Change*, 144(2):347–364.

Xia, H. and Carlin, B. P. (1998). Spatio-temporal models with errors in covariates: Mapping Ohio lung cancer mortality. *Statistics in Medicine*, 17(18):2025–2043.

Xiu, D. and Karniadakis, G. E. (2002). The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2):619–644.

Xu, L., Paterson, A. D., Turpin, W., and Xu, W. (2015). Assessment and selection of competing models for zero-inflated microbiome data. *PloS one*, 10(7).

Zhou, Y., Johansen, A. M., and Aston, J. A. (2016). Toward automatic model comparison: an adaptive sequential Monte Carlo approach. *Journal of Computational and Graphical Statistics*, 25(3):701–726.

Zhu, G., Li, X., Ma, J., Wang, Y., Liu, S., Huang, C., Zhang, K., and Hu, X. (2018). A new moving strategy for the sequential Monte Carlo approach in optimizing the hydrological model parameters. *Advances in Water Resources*, 114:164–179.

Vita

Seiyon Lee

Education

- Ph.D. Statistics, Pennsylvania State University, 2020.
  Thesis Title: *Computational Methods for Hierarchical Spatial Models and Ice Sheet Model Calibration* (Thesis advisor: Dr. Murali Haran, Thesis co-advisor: Dr. Klaus Keller)
- M.A. Applied Statistics, Columbia University, 2015
- B.A. International Studies, Johns Hopkins University, 2008

Publications

- Lee, B.S. and Haran, M. (2020+) Modeling and Computation for High-dimensional Zero-Inflated Spatial Data. *In preparation.*
- Lee, B.S. and Haran, M. (202+) PICAR: An Efficient Extendable Approach for Fitting Hierarchical Spatial Models. *Under review*
- Lee, B.S., Haran, M., Fuller, R.W., Pollard, D., and Keller, K. (2019+) A Fast Particle-Based Approach for Calibrating a 3-D Model of the Antarctic Ice Sheet. To appear in *The Annals of Applied Statistics*
- Lee, B. S., Haran, M., and Keller, K. (2017). Multidecadal Scale Detection Time for Potentially Increasing Atlantic Storm Surges in a Warming Climate. *Geophysical Research Letters*, 44, 10,617– 10,623.
- Oddo, P. C., Lee, B. S., Garner, G. G., Srikrishnan, V. , Reed, P. M., Forest, C. E. and Keller, K. (2017), Deep Uncertainties in Sea-Level Rise and Storm Surge Projections: Implications for Coastal Flood Risk Management. *Risk Analysis*.

Honors and Awards

- Honorable Mention, 2020 American Statistical Association Section on Statistics in the Environment (ENVR) Student Paper Competition
- Winner, 2019 J. Keith Ord Scholarship for Research in Spatial and Environmental Statistics
- Graduate Fellow, 2016 Jack and Eleanor Pettit Scholarship in Science
- Graduate Fellow, 2015 University Graduate Fellowship