

The Pennsylvania State University
The Graduate School

CLIMATE MODEL CALIBRATION USING HIGH-DIMENSIONAL
AND NON-GAUSSIAN SPATIAL DATA

A Dissertation in
Statistics
by
Won Chang

© 2014 Won Chang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2014

The dissertation of Won Chang was reviewed and approved* by the following:

Murali Haran
Associate Professor of Statistics
Dissertation Advisor, Chair of Committee

Klaus Keller
Associate Professor of Geoscience
Dissertation Co-Advisor

Bing Li
Professor of Statistics

Donald Richards
Professor of Statistics

Aleksandra Slavkovic
Associate Professor of Statistics
Chair of Graduate Program

*Signatures are on file in the Graduate School.

Abstract

This thesis focuses on statistical methods to calibrate complex computer models using high-dimensional spatial data sets. This work is motivated by important research problems in climate science where such computer models are frequently used. Climate models play a central role in generating projections of future climate. An important source of uncertainty about future projections from these models is due to uncertainty about input parameters that are key drivers of the resulting hindcasts and projections. Climate model calibration is a statistical framework for inferring the input parameters by combining information from climate model runs and observational data. When the data are in the form of high-dimensional spatial fields, climate model emulation (approximation) and calibration can pose significant modeling and computational challenges. The goal of this research is to develop new approaches to computer model calibration that are computationally efficient, accurate, and carefully account for uncertainties. The main contributions of this thesis are three-fold: (1) to develop a highly efficient reduced-dimensional climate model calibration approach that enables the use of high-dimensional spatial data; (2) to formulate a novel calibration method based on block composite likelihood and study the asymptotic properties of the resulting estimates for input parameters; and (3) to introduce a calibration framework that generalizes the existing method to the one-dimensional exponential family and formulate a reduced-dimensional approach that can efficiently handle the high-dimensional non-Gaussian spatial data. Our methods provide insights about current and future climate. In our first application we make projections of the North Atlantic Meridional Overturning Circulation (AMOC), an ocean circulation that transports heat from low- to high-latitude areas in the Atlantic and contributes to the mild climate in Northern and Western Europe. AMOC changes are projected to impact human and natural systems. We demonstrate that utilizing information from high-dimensional spatial data reduces parametric uncertainty and thus results in an AMOC projection with reduced uncertainties. In the second case study, we demonstrate an application of our approach for non-Gaussian spatial data to calibration of a Greenland ice sheet model and show that our approach can improve upon current methods for projections of sea level rise contributed by the Greenland ice sheet.

Table of Contents

List of Figures	vii
List of Tables	xiii
Acknowledgments	xiv
Chapter 1	
Introduction	1
1.1 Background and Motivating Climate Science Problem	2
1.2 Statistical Problem Definition and Challenges	4
1.3 Contributions	6
1.4 Thesis Organization	6
Chapter 2	
Gaussian Processes, Computer Models, and Large Spatial Data	8
2.1 Gaussian Processes, Spatial Modeling and Matérn Covariance	8
2.2 Computer Models	10
2.2.1 Computer Model Emulation	11
2.2.2 Computer Model Calibration	12
2.2.3 Computational Challenges	14
2.3 Basis Representation Approaches	16
2.3.1 Principal Component Analysis	17
2.3.2 Kernel Convolution	18
2.3.3 Limitations and Possible Solutions	18
2.4 Summary	19
Chapter 3	
Composite Likelihood for Spatial Data	20
3.1 Overview of Composite Likelihood for Large Spatial Data	21
3.2 Theoretical Issues	23
3.2.1 Asymptotic Properties of MLE	23

3.2.2	Asymptotic Properties of MCLE	26
3.3	Bayesian Inference Using Composite Likelihood	27
3.3.1	Asymptotic Properties of Composite Posterior Estimators	28
3.4	Posterior Sample Adjustment	31
3.5	Discussion	31

Chapter 4

	Fast Dimension-Reduced Climate Model Calibration and the Effect of Data Aggregation	33
4.1	Introduction	34
4.2	Data Description	37
4.3	Model Calibration Framework	38
4.3.1	Two-Stage Emulation and Calibration	40
4.3.2	Challenges with High-Dimensional Data	41
4.4	Model Calibration with High-Dimensional Spatial Data	41
4.4.1	Current Approaches	42
4.4.2	Reduced Dimensional Model Calibration	43
4.4.2.1	Computer Model Emulation	43
4.4.2.2	Computer Model Calibration	45
4.5	Implementation Details	49
4.6	Results	51
4.6.1	Computational Benefit	51
4.6.2	The Effect of Data Aggregation on Climate Model Calibration	52
4.7	Discussion	54
4.7.1	Caveats	54
4.7.2	Summary	55

Chapter 5

	Computer Model Calibration Using Block Composite Likelihood	64
5.1	Introduction	64
5.2	Calibration Using Gaussian Processes	66
5.3	Calibration with High-Dimensional Spatial Data	68
5.3.1	Challenges with High-Dimensional Spatial Data	69
5.3.2	Composite Likelihood for Model Calibration	70
5.4	Application to UVic ESCM Calibration	79
5.4.1	Simulated Examples	80
5.4.2	Calibration Using Observed Ocean Temperature Anomalies	82
5.5	Discussion	83
5.5.1	Caveats	83
5.5.2	Summary and Future Direction	84

Chapter 6	
Ice Sheet Model Calibration Using High-Dimensional Binary Spatial Data	90
6.1 Motivating Problem from Ice Sheet Model	91
6.2 Ice Sheet Model Calibration Using Ice Thickness Profile	93
6.3 Computer Model Emulation and Calibration Using Binary Spatial Data	94
6.3.1 Computer Model Emulation Using Binary Spatial Data	95
6.3.2 Computer Model Calibration Using Binary Spatial Data	96
6.3.3 Computational Challenges	96
6.4 Reduced-Dimension Approach for Binary Spatial Data	97
6.4.1 Dimension Reduction for Binary Model Output	98
6.4.2 Emulation Using Logistic PCs	100
6.4.3 Calibration Using Logistic PCs	101
6.5 Implementation Details and Results	102
6.5.1 Simulated Example	102
6.5.2 Calibration of an Ice Sheet Model Using Binary Ice Patterns	103
6.6 Discussion	106
Chapter 7	
Discussion	114
7.1 Summary and Contributions	114
7.2 Caveats and Potential Improvements	115
7.3 Avenues for Future Research	117
Appendix A	
Reduced-Dimensional Calibration.	120
A.1 Computational Cost and Storage Issue in Emulation Problem	120
A.2 Covariance Structure Implied by Separability Assumption	121
A.3 Diagnostics for Prediction Performance of PC Emulator	121
A.4 Computing Time Reduction by Parallel Computing	122
Appendix B	
Composite Likelihood Calibration	131
B.1 Computation of \mathbf{P}_n and \mathbf{Q}_n	131
Bibliography	135

List of Figures

2.1	A simulated example illustrating computer model emulation and calibration. (a) The mean of the Gaussian process emulator (dashed red line) interpolates model runs (black dots) and the 95% prediction interval (dotted red lines) shows the interpolation uncertainty. The observational data is given as a scalar (solid blue line) with a normally distributed measurement error (solid black curve on the right side of the plot). The dashed blue lines show 95% confidence interval for the observational data. (b) The posterior density is computed by the standard calibration approach that combines the emulator output and the observational data.	13
4.1	Plots of ocean temperature patterns averaged over 1955–2006. A UVic ESCM run (Srifer et al., 2012, the first row) and the observational data from World Ocean Atlas 2009 (Antonov et al., 2010; Locarnini et al., 2010, the second row). The left column shows the latitude-depth profiles of zonal mean, and the right column displays the longitude-latitude profiles of vertical mean. Note that the model run shown here is an example of 250 model runs described in Section 4.2.	58
4.2	Comparison of computational costs for the emulation step between the current approaches and the new approach. The green box near the bottom right corner shows computing times that are practical, ranging from one second to three months.	59

4.3	<p>Prior sensitivity test in the simulated example. Calibration of K_{bg} value based on (a) 1-D depth profile, (b) 2-D latitude-depth pattern (c) 3-D non-aggregated data. Each line represents posterior density from four different priors: $(b_v = 2, b_z = 2)$ (Prior 1, solid red line), $(b_v = 2, b_z = 100)$ (Prior 2, dashed red line), $(b_v = 100, b_z = 2)$ (Prior 3, solid blue line), and $(b_v = 100, b_z = 100)$ (Prior 4, dashed blue line). The solid vertical line represents the true value of K_{bg} in the synthetic truth. b_v and b_z are hyperparameters of κ_d and σ^2 respectively in (4.5).</p>	60
4.4	<p>Prior sensitivity test using observational data from the World Ocean Atlas 2009 (Antonov et al., 2010; Locarnini et al., 2010). Calibration results based on (a) 1-D depth profile, (b) 2-D latitude-depth pattern, (c) 3-D non-aggregated data. Each line represents posterior density from four different priors: $(b_v = 2, b_z = 2)$ (Prior 1, solid red line), $(b_v = 2, b_z = 100)$ (Prior 2, dashed red line), $(b_v = 100, b_z = 2)$ (Prior 3, solid blue line), and $(b_v = 100, b_z = 100)$ (Prior 4, dashed blue line), where b_v and b_z are hyperparameters of κ_d and σ^2 respectively in (4.5).</p>	61
4.5	<p>Combined posterior densities of K_{bg} from different prior specifications (lower left), the relationship between K_{bg} and projected AMOC change of the 2070 - 2099 mean from the 1970 - 1999 mean (upper left), and the resulting AMOC change projections (upper right) using 1-D (solid black line), 2-D (dashed red line) and 3-D (dotted blue line) data with their 95% credible intervals (bars at the right).</p>	62
4.6	<p>An example of our leave-10-percent out cross validation experiment for 2-D case (latitude-depth patterns). The first row shows the comparison between the raw original output and the raw emulated output. The second row shows the same comparison using detrended outputs, which are computed by subtracting the mean across the parameter settings at each location.</p>	63

5.1	Comparison between calibration results using i) the original likelihood without blocking (solid black curves), ii) the block composite likelihood without the variance adjustment (dashed red line), and iii) the block composite likelihood with the variance adjustment (dashed-dotted blue line). The vertical lines represent the assumed true value for our simulation, and the horizontal bars above show the 95% credible intervals. The results shown here are based on two different realizations (two left panels for Realization 1 and two right panels for Realization 2) from the same Gaussian process model. The posterior densities with the approximation for the block means (two lower panels) are reasonably close to the densities without the approximation (two upper panels) when the variance adjustment is applied.	86
5.2	Comparison of posterior densities between three simulated examples with different block numbers: $M = 10$ (solid black curve), $M = 30$ (dashed red curve), and $M = 50$ (dotted-dashed blue curve). The vertical line is the assumed true value for our simulated example and the horizontal bars above are 95% credible intervals. Posterior modes based on 30 and 50 blocks show slight biases, but the width of intervals does not show notable differences.	87
5.3	Comparison of posterior densities between three simulated examples with different assumed magnitudes of the discrepancies: $\kappa_d^* = 40000$ (solid black curve), $\kappa_d^* = 90000$ (dashed red curve), and $\kappa_d^* = 160000$ (dotted-dashed blue curve). The vertical line indicates the assumed true value, and the horizontal bars above show the 95% credible intervals. As the discrepancy grows, the densities become more dispersed but the posterior modes stay similar.	88
5.4	Posterior densities of the climate sensitivity calibrated based on the observational data from Levitus et al. (2012) using our composite likelihood approach. The adjusted posterior density (solid black curve) is notably more dispersed than the unadjusted one (dashed black curve), and the corresponding 95% credible intervals (horizontal bars above) for the adjusted posterior density is also much wider than the one for the unadjusted density.	89

6.1	Comparison of the estimated discrepancy patterns by the procedure described in 6.4.2 (left panels) and the true discrepancy patterns that are actually used in 6.5 (right panels). The upper panels show the results for one of the discrepancy patterns used in the simple example in 6.5.1. The other 19 cases with different realizations of the discrepancy process exhibit the similar results. The lower panels show the discrepancy patterns in the SICOPOLIS calibration problem in Section 6.5.2. In both cases the estimated discrepancies well capture the overall patterns of the true discrepancy patterns.	107
6.2	Example of 20 model runs out of total 100 model runs. The resolution of each image is 30×30 . The first parameter θ_1 controls the overall size of the grey pattern and the second parameter θ_2 changes both the shape and the size of the pattern.	108
6.3	The joint (upper panel) and the marginal (lower two panels) posterior densities of the input parameters based on synthetic observational data sets with 20 different realizations from the discrepancy process defined in 6.5.1. Most of the cases the resulting posterior densities recover the true input parameter setting (white dashed lines in the upper panel and black solid line in the lower panels) reasonably well. The marginal posterior distributions have significantly smaller variation comparing to the prior distribution, showing that the parametric uncertainty has been significantly reduced by calibration.	109
6.4	The output from the original model run chosen as the synthetic truth (left panel) and the synthetic observational data obtained by superimposing the discrepancy term described in Section 6.5.2. The synthetic observation has a notably smaller ice covered area.	110
6.5	An example of leave-10-percent-out cross-validation result to verify the prediction performance of our emulator based on logistic PCA. The result shown here is for model run #74 in Applegate et al. (2012). The graphical comparison shows that the emulated model output (right panel) predicts the original SICOPOLIS model output precisely. Similar results hold for the other model runs.	111
6.6	The pairwise joint posterior densities for the input parameters from the calibration results in Section 6.5.2. The white dashed lines show the input parameter settings for the synthetic truth. Only the marginal posterior density of the ice PDD factor peaks around the values for the synthetic truth. The marginal densities for the snow PDD factor and the basal sliding factor show bimodality, which requires further investigation.	112

6.7	The ice volume change projections from 2006 to 2011 generated by the SICOPOLIS Applegate et al. (2012) based on our calibration approach (solid red line), the 10% area filter that chooses model runs within 10% of the observed ice covered area (solid blue line), and no calibration (dashed green line). The parentheses on the horizontal bars represent 95% prediction interval and the vertical bars the ranges. Our approach yields a narrower 95% prediction interval than the filtering approach. The predictive distribution from our approach shows slight bias probably due to the biases in the posterior densities for the input parameters other than the ice PDD (see Figure 6.6).	113
A.1	Comparison of UVic ESCM grid (Weaver et al., 2001) and World Ocean Atlas grid (Antonov et al., 2010; Levitus et al., 2012), projected on latitude-depth space. The pattern embedded in each plot is the mean ocean temperature averaged over longitude (0-360°) and time (1955-2006). Note that the temperature pattern shown in (a) is an example of 250 model runs described in Section 2.	123
A.2	Comparison of UVic ESCM grid and World Ocean Atlas grid, projected on longitude-latitude space. For better visibility of the figures we display only the North Atlantic area, but the model output and observations cover the entire globe. The pattern embedded in each plot is the mean ocean temperature averaged over depth (0-3000m) and time (1955-2006). Note that the temperature pattern shown in (a) is an example of 250 model runs described in Section 4.2.	124
A.3	Comparison of emulation performance between the exponential covariance function and the squared exponential covariance function based on 50 different leave-10-percent-out cross validation scenarios. In each scenario, we randomly hold out 25 different 3-D model outputs and predict them using the remaining 225 model runs. The root mean squared errors (RMSe) are computed for each method, and the RMSe for the exponential covariance function is subtracted from the RMSe for the squared exponential function. The boxplot of differences in RMSe shows that squared exponential covariance function outperforms or shows similar performance to the exponential covariance function. Any positive RMSe corresponds to squared exponential covariance outperforming exponential covariance.	125

A.4	Scree plot showing the eigenvalues for the 3-D case. The magnitude of the eigenvalues decreases rapidly for the first 10 principal components, and hence the remaining eigenvalues are extremely small comparing to the leading eigenvalues.	126
A.5	Graphical diagnostics using standardized uncorrelated prediction errors. For the first four principal components, prediction errors using pivotal Cholesky decomposition are computed at five K_{bg} settings (0.1, 0.2, 0.3, 0.4, 0.5) while the other two parameters are fixed at $A_{sc} = 1.5$ and $C_s = 3.976$, resulting in 20 uncorrelated standardized prediction errors.	127
A.6	Posterior densities of K_{bg} based on 10 different random samples from 3-D locations. Each curve represents the posterior density from one sample with size $n = 1,300$	128
A.7	Spatial patterns for the first principal component basis vector. The latitude-depth pattern seems to show the contrast between the near the surface part and the deeper part of the ocean in the low latitude area caused by the effect of K_{bg} on diffusion (Kriegler, 2005). The latitude-longitude pattern seems to show the effect of K_{bg} on AMOC in the high-latitude North Atlantic, Western Boundary Currents (primarily Kuroshio current), and Eastern Boundary Currents (e.g., Peru current).	129
A.8	Comparison of posterior densities of K_{bg} computed by a) our PCA-based approach (black solid line) and b) assuming separable covariance structure (red dashed line). The results are based on 2D (latitude-depth) ocean temperature patterns.	130

List of Tables

- 5.1 Summary of notation used in Section 5.2 67
- 5.2 Summary of notation used for emulation model in Section 5.3 . . . 69
- 5.3 Summary of notation used for calibration model in Section 5.3 . . . 70

Acknowledgments

I am deeply grateful to my advisers, Murali Haran and Klaus Keller for their mentorship and guidance on my research and professional development. I feel very fortunate to have worked with such exemplary mentors who are always supportive, encouraging and inspiring. And I thank my collaborator Roman Olson for his excellent work and tremendous help in the Earth system model calibration research that has resulted in two manuscripts included in this thesis. I also thank my collaborator Patrick Applegate for providing me a wonderful collaboration opportunity and giving me a tremendous guidance in the ice sheet model calibration research. I feel that his contribution to my dissertation research is significant enough to recognize him as my co-advisor.

Chapter 4 and 5 are based on manuscripts co-authored by Murali Haran, Klaus Keller, and Roman Olson and published in the *Annals of Applied Statistics* and *Statistica Sinica* respectively. I am grateful for the editors, Tilmann Gneiting and Ming-Hui Chen, and other anonymous reviewers for their comments and suggestions that have greatly improved the manuscripts. I also thank Ryan Sriver for providing the climate model runs used in the application examples, and Stefano Castruccio, Rob Nicholas, and Ben Shaby for useful discussions. Chapter 6 is based on a manuscript in preparation co-authored by Patrick Applegate, Murali Haran, and Klaus Keller. I thank Ralf Greve for distributing his ice sheet model SICOPOLIS, used in the application example in this chapter, freely on the Web (<http://sicopolis.greveweb.net/>).

I would like to express my gratitude to all faculty members in the Department of Statistics at Penn State. In particular I thank Bing Li and Donald Richards for serving as my dissertation committee members and providing useful comments, and Debashis Ghosh and David Hunter for serving as references for my job applications. I also thank Dennis Lin and Naomi Altman for their advice on job interview preparation.

A special thanks to my family and friends. I would like to express appreciation to my beloved wife Jeonga Woo for her unwavering support. I also thank my fellow

students and friends including Daisy Philtron, Jason Philtron, Vishesh Karwa, Kwame Kankam, Jason Berstein, Josh Goldstein, Yawen Guan, Saksham Chandra, and Seonjin Kim who made my graduate life at Penn State very pleasant and memorable.

This dissertation research was partially supported by the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507.

Chapter 1

Introduction

Computer models are indispensable tools in modern science and engineering. They are very useful for studying complex phenomena in various fields including, for example, climate science, glaciology, weather forecasting, disease dynamics, ecology, mechanical engineering, industrial engineering, and hydrology. One of the important uses of computer models in the context of climate science is in generating projections for future climate; these projections are in turn important for studying the potential impacts of climate change to inform decision making. The projections from the computer models are usually uncertain due to uncertainty in input parameters that characterize the behavior of computer models. Therefore learning about these parameters from observations and accurately quantifying uncertainties about these parameters as well as the resulting projections is an important research problem. This thesis develops and studies statistical approaches for parametric uncertainty quantification for computer models in climate science. Increasingly, both the output from these climate models as well as the observations are in the form of large spatial data sets and these spatial data may be Gaussian or non-Gaussian. This presents a number of challenges both in terms of developing appropriate models for such data as well as solving the considerable computational challenges imposed by the size of the data set. This work describes new models and computationally efficient methods that are able to take full advantage of the information available in these large spatial data sets while still rigorously accounting for multiple and complicated sources of uncertainty. The statistical methodology presented here is motivated by specific climate science research problems but the

new methods developed are generally applicable, in principle, to any computer model where the output is in the form of large spatial data.

1.1 Background and Motivating Climate Science Problem

There is a growing consensus in the scientific community that the greenhouse gas emissions caused by human activity is changing the Earth's climate system and this in turn may increase climate risk (Alexander et al., 2013). Assessing the possibility and the impact of such events is important in climate risk analysis, and generating projections from climate models is one of the main avenues to approach this problem (Schneider, 2001). The projections from the climate models are in general considerably uncertain due in part to uncertainty about climate model ("input") parameters. The climate parameters may represent physical processes that are not resolved at the relevant spatio temporal scales in the climate models (cf. Goes et al., 2010; Knutti et al., 2002; Knutti and Hegerl, 2008). For example they might represent the processes that are important but operating below the model scale, or not understood well enough to be explicitly described in the model equations. While quantifying and reducing the parametric uncertainty is crucial for generating more informative climate projections, inferring these parameters is nontrivial since they are not typically directly measurable from climate observations. There has been increasing interest in procedures for indirectly inferring climate parameters by combining observable variables in climate models and observational data from real world climate (see, e.g. Bhat et al., 2012; Forest et al., 2002, 2008; Olson et al., 2012; Sansó and Forest, 2009). Roughly speaking, these are all statistical approaches for estimating climate parameter values that produce the most compatible climate processes to the observed climate processes, while accounting for various sources of errors. The idea is therefore to use these parameter estimates to produce realistic future climate projections along with uncertainties about the projections.

I briefly discuss here two examples of climate projection problems that motivate the methodological development in this thesis. The first example is a projection of

the Atlantic meridional overturning circulation (AMOC) (Rahmstorf, 1997). The AMOC is a part of the global scale thermohaline circulation in the North Atlantic ocean driven by temperature and salinity gradients. The circulation transports heat from low- to high-latitude areas in the Atlantic and contributes to the mild climate in Northern and Western Europe. AMOC changes are projected to impact human and natural systems (cf. Alley et al., 2007; Keller et al., 2005, 2007). The objective here is to generate realistic projections for AMOC using a climate model. The projections for AMOC are affected by the strength of vertical mixing in the ocean (Wunsch and Ferrari, 2004). Since the ocean mixing is affected by processes that operate below the typical model gridbox scale, it needs to be parameterized. One approach to this end is to introduce a parameter such as the background vertical diffusivity (K_{bg}) (cf. Weaver et al., 2001; Schmittner et al., 2009; Goes et al., 2010). The value of K_{bg} also affects the ocean temperatures, which are observable from the real ocean, and hence it is possible to estimate K_{bg} by comparing the ocean temperature output from the climate model with oceanographic measurements.

The second example is projecting changes in the Greenland ice sheet. Greenland ice volume loss is potentially an important contributor to global sea-level rise that may in turn result in high-impact climate events such as storm surges and coastal flooding (Bamber et al., 2001; Bamber and Aspinall, 2013; Alexander et al., 2013; Lemke et al., 2007). Various ice sheet models have been proposed, but ice volume change projections from these models are deeply uncertain due to various unknown input parameters (cf. Applegate et al., 2012). A warmer climate can accelerate ice melting through increased atmospheric and ocean temperature, but also increase ice-sheet growth by increased precipitation. The overall volume change of the Greenland ice sheet due to the warmer climate depends on the values of parameters that control the processes in the ice sheet such as ice flow, basal sliding, surface melting, and sub-shelf melting. As in the AMOC projection task these parameters are not directly observable and therefore need to be estimated by comparing models output and observations of key constraints such as the modern ice-sheet profile.

There are several important sources of uncertainty surrounding these problems including parametric uncertainty, model structure uncertainty, code uncertainty, and observation error (Kennedy and O'Hagan, 2001; Walker et al., 2003). As al-

ready discussed above, parametric uncertainty refers to the uncertainty caused by inability to determine exact values of input parameters. Model structure uncertainty is the uncertainty due to inadequate model representation of the real system that cannot be removed by parameter specifications (Kennedy and O’Hagan, 2001; Walker et al., 2003). Code uncertainty is the uncertainty caused by the fact that model output is usually available only at a limited number of parameters settings. This occurs because the climate models are computationally expensive due to their complexity and need for long-term spin-up, and as a result each of the model runs typically takes multiple hours to days or even months. Observation error is measurement uncertainty in obtaining observations that are being compared by climate model output. Quantifying parametric uncertainty while accounting for other sources of uncertainty is challenging. I outline some of the challenges below.

1.2 Statistical Problem Definition and Challenges

The main goal of this thesis is to formulate statistical methods for inferring input parameters for climate or ice sheet models. The procedure of inferring input parameters by combining model output and observational data is referred to as computer model calibration. The methodologies developed in this work are built upon the statistical calibration framework proposed by Kennedy and O’Hagan (2001). The approach enables estimation of parameters using a probability model that relates the input parameters to the observational data through the computer model output while allowing for systematic and observational errors. Calibrating the climate model and the ice model described in the previous section poses various challenges: The first challenge is related to code uncertainty. The calibration approach requires obtaining model output at a large number of parameter settings but the number of available model runs is limited in these problems due to the high computational cost for each model run. One standard approach is to substitute the computer model by a fast approximate model (called emulator), constructed by interpolating between model output at a set of design points in the input parameter space (Sacks et al., 1989). In our problem building an emu-

lator poses considerable computational and inferential challenges since the model output is in the form of high-dimensional spatial data. Another challenge lies in comparing the emulator output and the observational data to account for model structure uncertainty in calibration. The error between the model output and the observational data is usually a complicated high-dimensional spatial process. This causes not only the computational issues related to dealing with large spatial data, but also the identifiability issues in estimating the input parameters. The calibration framework using a Gaussian process emulator has been successfully applied for climate model calibration problems (cf. for example Sansó and Forest, 2009; Olson et al., 2012), but these approaches are based on heavily aggregated data and hence are silent on aspects of key high-dimensional spatial processes. A natural approach to the challenge of computer model emulation-calibration with high-dimensional spatial data is to build upon some of the approaches developed in the spatial statistics literature. Modeling high-dimensional spatial data has gained a lot of recent attention in the geostatistics community in the context of interpolating observational data (Higdon, 1998; Banerjee et al., 2008; Cressie and Johannesson, 2008; Fuentes, 2007; Vecchia, 1988; Stein et al., 2004; Caragea and Smith, 2006; Eidsvik et al., 2013; Furrer et al., 2006; Lindgren et al., 2011; Simpson et al., 2012). However, there is relatively little work on emulation-calibration with high-dimensional spatial data. The calibration method described by Bhat et al. (2012) can partially address this issue but the approach still does not scale to the problems and data set sizes we consider. Other calibration approaches for high-dimensional data developed in different contexts (Bayarri et al., 2007; Higdon et al., 2008) are applicable but also have some limitations that prevent them from being applicable to high-dimensional spatial data from climate model output and observations (see Chapter 4 for more discussion). Furthermore, model output and observational data are often in the form of spatially correlated *non-Gaussian* data. For instance, this arises in the context of the ice model calibration problem since important model output variables such as ice thickness and ice shapes are in the form of non-Gaussian spatial processes. These kind of data pose non-trivial modeling and computational challenges, due to the non-Gaussianity of the data, the size of the data, as well as the need to account for errors, in particular the data-computer model discrepancies. To my knowledge calibration using non-Gaussian

spatial data is a largely unexplored area.

1.3 Contributions

This thesis makes the following main contributions to statistical methodology and climate science:

1. A new computationally expedient and nearly automated emulation approach that provides an accurate approximation to climate models.
2. A new computationally efficient calibration framework based on dimension reduction that can deal with spatial data with more than tens of thousands of spatial locations while still permitting a flexible error structure.
3. A formal investigation of the effect of data aggregation on climate model parameter inference. The results indicate that data aggregation causes considerable information loss.
4. The application of composite likelihood methods to the emulation and calibration problem, along with a study of the asymptotic properties of the input parameter estimates based on block composite likelihood. To my knowledge, this is the first application of composite likelihood to this problem.
5. Formulating an emulation and calibration approach to model output and observations in the form of non-Gaussian spatial data for the first time. The method can efficiently deal with high-dimensional non-Gaussian spatial datasets with more than tens of thousands of spatial locations.

1.4 Thesis Organization

The remainder of this thesis is organized as follows. In Chapter 2 I review the basics of Gaussian process modeling and inference for spatial data, computer model calibration using a Gaussian process emulator and reduced-dimension approaches for modeling large spatial data. Chapter 3 provides an overview of composite likelihood for spatial data. I summarize the basics of the composite likelihood

framework and introduce different types of composite likelihood for spatial modeling in Section 3.1. In Section 3.2 I examine the theoretical properties of the estimators based on composite likelihood for spatial data and in Section 3.3 and 3.4 I study the use of composite likelihood in the Bayesian framework. Chapter 4 describes a computationally efficient reduced-dimension approach that enables the application of calibration methods to large spatial data sets. Using this approach I investigate the effect of data aggregation on inference of the vertical ocean diffusivity and show that using aggregated data leads to information loss and hence using high-dimensional spatial data leads to a better calibration result. This work, joint with Murali Haran, Roman Olson and Klaus Keller, has been published in the *Annals of Applied Statistics*. In Chapter 5 I introduce an alternative approach to climate model calibration that uses block composite likelihood. I derive the asymptotic properties of the estimators for input parameters based on the results described in Chapter 3 and use them to formulate a method that accounts for the information loss caused by substituting the likelihood function with a composite likelihood function. This work, joint with Murali Haran, Roman Olson and Klaus Keller, has been accepted for publication in *Statistica Sinica*. In Chapter 6 I turn my attention to calibration using non-Gaussian data motivated by calibration of a computer model for the Greenland ice sheet. I propose a reduced-dimension calibration approach for binary spatial data, which can be easily generalized to the one-parameter exponential family distributions. This ongoing work is joint with Murali Haran, Patrick Applegate and Klaus Keller. Finally in Chapter 7 I summarize the contributions of this thesis and discuss directions for future work.

Gaussian Processes, Computer Models, and Large Spatial Data

In this chapter we overview the basics of computer model emulation and calibration. The computer model emulation approach that we use here heavily relies upon the Gaussian process framework, and hence we begin with a description of Gaussian process basics. We discuss the use of Gaussian processes for modeling spatial processes and introduce the Matérn covariance function that plays a crucial role in spatial modeling. We then describe the standard approach to computer model emulation and calibration and explain computational challenges posed by large spatial datasets. We conclude the chapter by outlining dimension reduction approaches for modeling large spatial data which we apply to computer model calibration in Chapter 4 to solve the computational issues.

2.1 Gaussian Processes, Spatial Modeling and Matérn Covariance

Let \mathcal{S} be an arbitrary space with a distance function d . A Gaussian process $\{Z(\mathbf{s}); \mathbf{s} \in \mathcal{S}\}$ is continuous infinite dimensional stochastic process whose realizations $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$ at finite points $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathcal{S}$ follow a multivariate normal distribution

$$\mathbf{Z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

with an n -dimensional mean vector $\boldsymbol{\mu}$ and an $n \times n$ dimensional covariance matrix $\boldsymbol{\Sigma}$. The mean vector is often defined as a linear function $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, with an $n \times b$ covariate matrix \mathbf{X} and a b -dimensional coefficient vector $\boldsymbol{\beta}$. The covariance matrix is defined as

$$\{\boldsymbol{\Sigma}\}_{i,j} = C(\mathbf{s}_i, \mathbf{s}_j),$$

where the covariance function $C(\cdot, \cdot)$ satisfies the symmetry $C(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{s}_j, \mathbf{s}_i)$ and the positive definiteness $\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{s}_j, \mathbf{s}_i) > 0$ for any $a_1 \dots a_n$. If the covariance function depends on \mathbf{s}_i and \mathbf{s}_j only through the distance between the two points, the Gaussian process is called isotropic.

One useful class of Gaussian processes arises when the function C is a decreasing function of the distance between two different points. For given realizations $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$ the predictive process $\{Z(\mathbf{s}^*) | \mathbf{Z}; \mathbf{s}^* \in \mathcal{S}\}$ acts as a stochastic interpolator between the given points. The mean of the predictive process interpolates the points $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$ and its conditional covariance quantifies the uncertainty regarding the interpolation. Hence Gaussian processes are particularly useful in estimating an unknown function when the values of the function are available at only limited number of points.

An important application of this type of Gaussian process model is interpolation of spatial data where the space \mathcal{S} corresponds to the spatial region of interest and $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$ are observations at a set of spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ in that region. The distance function d is defined as a physical distance between two points such as the Euclidean or the geodesic distance. Often data of this form are called geostatistical data and interpolation between the observed data in this case is called “kriging”, named after in honor of a South African miner D. G. Krige (Mathéron, 1963). The approach was independently introduced by Krige (1951) and Matérn (1960) and since then Kriging has been a central tool in modeling spatial processes (cf. Cressie, 1993; Stein, 1999; Gelfand et al., 2010). In this type of application the covariates in \mathbf{X} consist of the spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ and other spatial processes that are expected to influence the process of interest. For example, if we are interested in spatial interpolation of precipitations measured at different weather stations, the covariates may include measured temperatures and humidities as well as the locations of the stations.

There are many possible choices for the covariance function C for the Gaussian

process interpolation, but we will focus on the Matérn covariance function that is the standard choice in geostatistics and has popular covariance functions such as the exponential and the squared exponential covariance functions as special cases. We refer to Whittle (1954) and Stein (1999) for some theoretical background for choosing the Matérn covariance function over other possibilities. For any given spatial locations $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$ the Matérn covariance function is defined as

$$C(\mathbf{s}, \mathbf{s}') = \kappa \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\sqrt{2\nu\phi}d(\mathbf{s}, \mathbf{s}') \right)^\nu K_\nu \left(\sqrt{2\nu\phi}d(\mathbf{s}, \mathbf{s}') \right),$$

where $\kappa > 0$ is the sill parameter that determines the magnitude of the covariance, $\phi > 0$ is the range parameter that specifies how the correlation decays with distance, and $\nu > 0$ is the smooth parameter that controls the squared mean differentiability of the sample paths (see, e.g. Stein, 1999). The function K_ν is the modified Bessel function of the second kind with order ν . For $\nu = 0.5$, the Matérn covariance function is reduced to the exponential covariance function. For $\nu \rightarrow \infty$, the Matérn covariance function converges to the squared exponential covariance function.

2.2 Computer Models

Computer models are deterministic models designed to simulate real world processes. They are typically computer codes that describe complex partial differential equations that represent fundamental laws and parameterizations of the system of interest. Starting from an initial condition, we obtain output from these models by numerically integrating the differential equations. A computer model is deterministic in the sense that given the same input parameter values and initial conditions, it always yields the same output. However, in most cases the output is largely controlled by input parameter values and hence output from computer models are subject to parametric uncertainties. In this section we introduce the computer model calibration framework, a statistical method to quantify and reduce the parametric uncertainty, and discuss the computational challenges posed by using large spatial data in this approach. In particular we focus on the two stage approach that approximates the computer model with a Gaussian process

emulator as the first step and infers the input parameters by combining the emulator and observational data as the second step (Bayarri et al., 2007; Bhat et al., 2010, 2012).

2.2.1 Computer Model Emulation

Computer models tend to have high computational costs since they are designed to represent complex processes such as the Earth’s climate system. Each run from these models may take a few hours to several days and therefore the number of available model runs is usually limited. This poses nontrivial challenges for computer model calibration because ideally we would need to run the computer model repeatedly at each (input) parameter setting of interest in order to understand how the computer model varies with the parameter. Here we resolve this challenge by constructing an “emulator”, a statistical model that serves as a fast approximation to the computer model. For ease of exposition here we focus on the case where the computer model yields a scalar outcome for each input parameter setting.

The first step of this approach is obtaining model output at a set of design points in the input parameter space $\Theta \subset R^d$ for an integer $d \geq 1$. We denote the design points for the input parameter settings by $\theta_1, \dots, \theta_p \in \Theta$ where p is the number of design points. Let $Y(\theta)$ be the model outcome at a parameter setting θ , then the output for the design points can be written as $\mathbf{Y} = (Y(\theta_1), \dots, Y(\theta_p))^T$. We build an emulator by interpolating these model runs using a Gaussian process in the parameter space Θ (Sacks et al., 1989) with a covariance function C with parameter vector ξ_y . The probability model for the existing model output is given by

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\xi_y)),$$

with the $p \times b$ covariate matrix \mathbf{X} , a b -dimensional coefficient vector $\boldsymbol{\beta}$, and the $p \times p$ covariance matrix $\boldsymbol{\Sigma}(\xi_y)$ given by

$$\{\boldsymbol{\Sigma}(\xi_y)\}_{ij} = C(\theta_i, \theta_j; \xi_y).$$

The covariates in \mathbf{X} are typically defined as the input parameter settings $\theta_1, \dots, \theta_p$ or set to be $\mathbf{0}$ since the trend induced by the covariance function C is often flexible

enough to capture the trend in \mathbf{Y} . We fit the Gaussian process by finding the maximum likelihood estimator (MLE) of $\boldsymbol{\xi}_y$, denoted by $\hat{\boldsymbol{\xi}}_y$.

The process for new model output $Y(\boldsymbol{\theta})$ at any untried parameter setting $\boldsymbol{\theta}$ is given by the conditional normal distribution of $Y(\boldsymbol{\theta})$ conditioning on the existing output \mathbf{Y} . We call this predictive process of $Y(\boldsymbol{\theta})$ an “emulator process” and denote it by $\eta(\boldsymbol{\theta}, \mathbf{Y})$. In general the emulator process based on a Gaussian process well captures a nonlinear trend of computer model output in $Y(\boldsymbol{\theta})$ and therefore provide a good approximation to the computer model. Moreover, the covariance of $\eta(\boldsymbol{\theta}, \mathbf{Y})$ informs us about the interpolation uncertainty. See Figure 2.1 (a) for an illustrative example of using a Gaussian process emulator for approximating a non-linear function.

2.2.2 Computer Model Calibration

Computer model calibration is a statistical procedure to infer input parameters by comparing emulator process $\eta(\boldsymbol{\theta}, \mathbf{Y})$ and observational data. While the focus of this procedure is on the parametric uncertainty, there are at least two other important sources of uncertainty that can affect the results: model-observation discrepancy and observational error. The model-observation discrepancy is the systematic error between the computer model and the real world process induced by misspecification or simplification of the physical process in the computer model. The observational error is a non-systematic error involving the measurement process. Sound statistical inference for the input parameters requires that we account for these sources of uncertainty.

Let Z be the observation data from the real world process that corresponds to the computer model output $Y(\boldsymbol{\theta})$. We set up the following probability model that incorporates all relevant sources of uncertainty:

$$Z = Y(\boldsymbol{\theta}^*) + \delta + \epsilon,$$

where $\boldsymbol{\theta}^*$ is the vector of input parameter values that best fit the data (Bayarri et al., 2007), $\delta \sim N(\mathbf{0}, \sigma_d^2)$ is the model-observation discrepancy with $\sigma_d^2 > 0$, and $\epsilon \sim N(\mathbf{0}, \sigma^2)$ is the observational error with $\sigma^2 > 0$.

Inferring the parameters in the above model requires repeated evaluation of

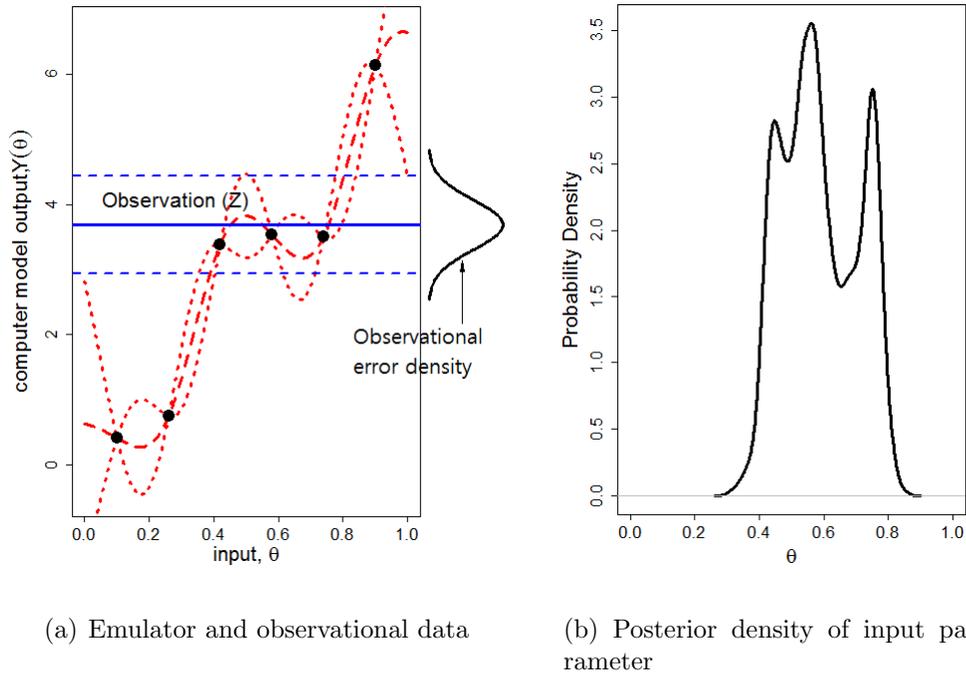


Figure 2.1: A simulated example illustrating computer model emulation and calibration. (a) The mean of the Gaussian process emulator (dashed red line) interpolates model runs (black dots) and the 95% prediction interval (dotted red lines) shows the interpolation uncertainty. The observational data is given as a scalar (solid blue line) with a normally distributed measurement error (solid black curve on the right side of the plot). The dashed blue lines show 95% confidence interval for the observational data. (b) The posterior density is computed by the standard calibration approach that combines the emulator output and the observational data.

the computer model $Y(\boldsymbol{\theta})$ for many different values of $\boldsymbol{\theta}$. This is not feasible as we discussed above, and we substitute the original computer model $Y(\boldsymbol{\theta}^*)$ with $\eta(\boldsymbol{\theta}^*, \mathbf{Y})$ and rewrite the probability model as

$$Z = \eta(\boldsymbol{\theta}^*, \mathbf{Y}) + \delta + \epsilon.$$

The corresponding likelihood function is straightforwardly given by the distributions of $\eta(\boldsymbol{\theta}^*, \mathbf{Y})$, δ , and ϵ . Here Bayesian inference is a more natural approach to this problem since the parameters are not identifiable without prior information; a Bayesian framework allows us to easily incorporate scientific knowledge about

these parameters. Based on the likelihood function for the probability model and some prior specifications for the parameters $\boldsymbol{\theta}^*$, σ_d^2 , and σ^2 , we can find the posterior density of $\boldsymbol{\theta}^*$ given \mathbf{Y} and Z . We can easily estimate the posterior distribution of the parameters by standard Markov Chain Monte Carlo (MCMC). Figure 2.1 shows a simple calibration example.

2.2.3 Computational Challenges

Computer models operating on large-scale and high-resolution spatial data sets have become common in many scientific fields including climate science. Climate models, for example, are designed to represent the Earth’s climate system and their outputs are typically spatial patterns of various climate variables such as the surface air temperature. Likewise, high-resolution observational data are also commonly available in many fields due to the development of observational instruments and data storage technology. In climate science high-resolution observational datasets are collected from various sources such as remote sensing.

Here we present the framework for computer model calibration using high-dimensional spatial data where “high-dimensional” here refers many spatial locations, and explain the computational challenges posed by using such datasets in calibration. Instead of a scalar output, we now have an n -dimensional vector of spatial patterns for each model run. We denote each model run at the parameter setting $\boldsymbol{\theta}$ by $\mathbf{Y}(\boldsymbol{\theta}) = (Y(\boldsymbol{\theta}, \mathbf{s}_1), \dots, Y(\boldsymbol{\theta}, \mathbf{s}_n))^T$ where $Y(\boldsymbol{\theta}, \mathbf{s})$ is the model output at an input parameter setting $\boldsymbol{\theta}$ and a spatial location $\mathbf{s} \in \mathcal{S}$. The spatial field \mathcal{S} represents the area that the computer model operates in. We have model runs $\mathbf{Y}(\boldsymbol{\theta}_1), \dots, \mathbf{Y}(\boldsymbol{\theta}_p)$ at p design points $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p \in \Theta$ as above and we denote the collection of these model runs by $\mathbf{Y} = (Y(\boldsymbol{\theta}_1, \mathbf{s}_1), \dots, Y(\boldsymbol{\theta}_1, \mathbf{s}_n), Y(\boldsymbol{\theta}_2, \mathbf{s}_1), \dots, Y(\boldsymbol{\theta}_p, \mathbf{s}_n))^T$. We let \mathbf{Z} be the vector of n -dimensional observational data that correspond to the model output $\mathbf{Y}(\boldsymbol{\theta})$.

We emulate the computer model output using the following Gaussian process model:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_y)), \quad (2.1)$$

where \mathbf{X} is a $np \times b$ matrix containing covariates and $\boldsymbol{\Sigma}(\boldsymbol{\xi}_y)$ is the $np \times np$ covariance matrix. The covariates in \mathbf{X} are functions of spatial locations and parameter

values that are used to define the covariance matrix $\Sigma(\boldsymbol{\xi}_y)$. The covariance matrix is defined so that the Gaussian process interpolates the model output both in the spatial field and the parameter space. The vector $\boldsymbol{\xi}_y$ contains the covariance parameters used to define $\Sigma(\boldsymbol{\xi}_y)$. As in the scalar output case, we fit the Gaussian process by finding the MLE of $\boldsymbol{\xi}_y$, denoted by $\hat{\boldsymbol{\xi}}_y$. The resulting emulator process $\boldsymbol{\eta}(\boldsymbol{\theta}^*, \mathbf{Y})$ is the n -dimensional predictive process for an untried parameter setting $\boldsymbol{\theta}^*$ given the existing model runs in \mathbf{Y} .

We calibrate the computer model using n -dimensional spatial observations \mathbf{Z} . We define the calibration model as

$$\mathbf{Z} = \boldsymbol{\eta}(\boldsymbol{\theta}^*, \mathbf{Y}) + \boldsymbol{\delta} + \boldsymbol{\epsilon}, \quad (2.2)$$

where $\boldsymbol{\delta} \sim N(\mathbf{0}, \Sigma(\boldsymbol{\xi}_d))$ is an n -dimensional Gaussian process with a vector of parameters $\boldsymbol{\xi}_d$ for the model-observation discrepancy and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ is the observational error process. Similar to the univariate case described above the input parameter $\boldsymbol{\theta}^*$ and the discrepancy parameter $\boldsymbol{\xi}_d$ are confounded and prior information is necessary for identifiability. With standard prior specifications for $\boldsymbol{\theta}^*$, $\boldsymbol{\xi}_d$, and σ^2 we define the posterior density for these parameters and infer them through MCMC.

In both emulation and calibration we face computational challenges when the size of spatial field n is large. For emulation stage each evaluation of the likelihood function corresponding to the probability model (2.1) requires storing the $np \times np$ covariance matrix and computing the inverse and the determinant of it. This requires storage space of $8n^2p^2$ bytes and computation that scales as $\mathcal{O}(p^3n^3)$, and both of them become quickly infeasible in a typical computation environment as np grows beyond tens of thousands. A few methods are proposed to overcome this difficulty (Bayarri et al., 2007; Bhat et al., 2012; Higdon et al., 2008), but these methods cannot fully resolve the computational issues when dealing with high-dimensional spatial data with tens of thousands of locations with many missing values, which are often the case for model output and observations in climate science (see Chapter 4 for more detailed discussion). We have similar challenges for calibration as well for dealing with the $n \times n$ covariance matrix and various approaches are studied to mitigate the challenges in modeling and interpolating

large spatial processes including basis representation (Higdon, 1998; Cressie and Johannesson, 2008; Banerjee et al., 2008; Nychka et al., 2013), composite likelihood (Besag, 1975, 1977; Caragea and Smith, 2006; Eidsvik et al., 2013; Heagerty and Lele, 1998; Stein et al., 2004; Vecchia, 1988), Whittle likelihood (Fuentes, 2007), Markov random fields (Lindgren et al., 2011; Simpson et al., 2012), and covariance tapering (Furrer et al., 2006; Kaufman et al., 2008; Shaby and Ruppert, 2012). Many of these methods can motivate formulation of computationally feasible emulation and calibration approaches. For the remainder of this chapter and in the next chapter we overview various approaches to mitigate the computational challenges in dealing with high-dimensional spatial data that motivate the calibration frameworks in Chapter 4 and 5.

2.3 Basis Representation Approaches

Modeling spatial data using Gaussian processes involves likelihood evaluation that requires Cholesky decomposition of $n \times n$ covariance matrix that scales as $\mathcal{O}(n^3)$ and storing the covariance matrix that needs memory allocation of $8n^2$ bytes. The basis representation approaches, also known as the low-rank approximation, solve the computational challenges by using spatial basis to represent the spatial data in the following form:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{\omega} + \boldsymbol{\epsilon},$$

where \mathbf{K} is an $n \times J$ matrix containing spatial basis with $J \ll n$, $\boldsymbol{\omega} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_w)$ is a J -dimensional vector of the coefficients with a $J \times J$ covariance matrix $\boldsymbol{\Sigma}_w$, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}_n)$ is a vector for the misrepresentation error. This above representation is equivalent to

$$\mathbf{Z} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{K}\boldsymbol{\Sigma}_w\mathbf{K}^T + \sigma^2\mathbf{I}_n),$$

and the inverse and the determinant of the covariance matrix are

$$\begin{aligned} (\mathbf{K}\boldsymbol{\Sigma}_w\mathbf{K}^T + \sigma^2\mathbf{I}_n)^{-1} &= \sigma^2\mathbf{I}_n - \sigma^{-2}\mathbf{K}(\boldsymbol{\Sigma}_w^{-1} - \mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\sigma^{-2}, \\ |\mathbf{K}\boldsymbol{\Sigma}_w\mathbf{K}^T + \sigma^2\mathbf{I}_n| &= n |\boldsymbol{\Sigma}_w - \sigma^{-2}\mathbf{K}^T\mathbf{K}| \sigma^2, \end{aligned}$$

respectively by Sherman-Morrison-Woodbury identity (Woodbury, 1950) and the Sylvester’s matrix determinant lemma (Harville, 2008). Both computations scales as $O(Jn^2)$ and this makes evaluation of the likelihood function computationally much more tractable for a moderately large n comparing to the likelihood computation for the original calibration model defined in 2.2. Moreover, we only need to store $n \times J$ matrix that requires memory space of only $8n^2J^2$ bytes.

Various approaches are proposed with different specifications for \mathbf{K} and ω . In this subsection we focus on the principal component analysis (PCA) and the kernel convolution that are used for formulating a reduced-dimension approach in Chapter 4. We briefly discuss the limitation of this approach and a possible solution towards the end.

2.3.1 Principal Component Analysis

The PCA approach uses the leading $J \ll n$ PCs for dimension reduction. We define basis matrix \mathbf{K} as the J leading eigenvectors of the covariance matrix Σ_z of \mathbf{Z} scaled by their corresponding eigenvalues. Let $\lambda_1, \dots, \lambda_J$ be the J leading eigenvalues and $\mathbf{e}_1, \dots, \mathbf{e}_J$ be their corresponding eigenvectors, then the basis matrix is given by $\mathbf{K} = (\sqrt{\lambda_1}\mathbf{e}_1, \dots, \sqrt{\lambda_J}\mathbf{e}_J)$. Letting $\omega \sim N(\mathbf{0}, \mathbf{I}_J)$, we obtain the approximation $\Sigma_z = \mathbf{K}\mathbf{K}^T + \sigma^2\mathbf{I}_n$. If an enough number of repeated observations on \mathbf{Z} are available, we can find \mathbf{K} from the sample covariance or, equivalently, using singular value decomposition (SVD). However, it is often the case that only one realization from \mathbf{Z} is available and in such case the covariance $\Sigma(\xi_z)$ needs to be specified with a covariance function and a set of covariance parameters ξ_z . This poses additional computational challenges since for each new value of ξ_z the likelihood evaluation involves eigendecomposition of $\Sigma(\xi_z)$, which scales as $\mathcal{O}(n^3)$. One approach to avoid this challenge is using the random projection approach described in Banerjee et al. (2013). The PCA approach yields the best likelihood approximation among the basis representation approaches with the same dimensionality J in terms of the Kullback-Leibler divergence (Stein, 2014).

2.3.2 Kernel Convolution

Kernel convolution was originally proposed by Higdon (1998) for mitigating the computational challenge of high-dimensional spatial data and, later, for allowing for more flexible non-stationary and non-isotropic covariances (Higdon, 1998; Higdon et al., 1999; Higdon, 2002). As the name suggests, the kernel convolution approach uses various kernel functions to define the columns of \mathbf{K} . Each column of \mathbf{K} is computed by kernel functions centered at J different locations $\mathbf{u}_1, \dots, \mathbf{u}_J \in \mathcal{S}$ (called ‘knots’) and $\boldsymbol{\omega}$ is a $J \times 1$ random vector where each of its elements corresponds to the Gaussian random fields at each of the knots. Each element of \mathbf{K} is defined by

$$k_j(\mathbf{s}_i) = C(\mathbf{s}_i, \mathbf{u}_j; \boldsymbol{\xi}_z), \quad (2.3)$$

where $C(\cdot, \cdot; \boldsymbol{\xi}_z)$ is a spatial covariance function between two points with a parameter $\boldsymbol{\xi}_z$. Higdon (1998) proposes to use a sparse and fixed covariance for $\boldsymbol{\omega}$ such that $\boldsymbol{\Sigma}_w = \mathbf{I}_J$ and estimate $\boldsymbol{\xi}_z$ for model fitting. Banerjee et al. (2008) defines the basis matrix as

$$\mathbf{K} = \boldsymbol{\Sigma}_{z, \boldsymbol{\omega}}(\boldsymbol{\xi}_z) \boldsymbol{\Sigma}_w(\boldsymbol{\xi}_z)^{-1},$$

where $\boldsymbol{\Sigma}_{z, \boldsymbol{\omega}}(\boldsymbol{\xi}_z)$ is the $n \times J$ covariance matrix between $\mathbf{s}_1, \dots, \mathbf{s}_n$ and $\mathbf{u}_1, \dots, \mathbf{u}_J$, and $\boldsymbol{\Sigma}_w(\boldsymbol{\xi}_z)$ is the $J \times J$ covariance matrix for $\mathbf{s}_1, \dots, \mathbf{s}_n$ determined by the underlying covariance function $C(\cdot, \cdot; \boldsymbol{\xi}_z)$. For these two approaches estimating $\boldsymbol{\xi}_z$ requires repeated computation of $\mathbf{K}^T \mathbf{K}$, which scales as $\mathcal{O}(Jn^2)$ and therefore may not be feasible for large n . To avoid this limitation Cressie and Johannesson (2008) fixes the parameters $\boldsymbol{\xi}_z$ and estimate the covariance of the coefficients $\boldsymbol{\Sigma}_w$ for model fitting. To prevent having a too restrictive representation, they use multi-resolutional basis computed by wavelets.

2.3.3 Limitations and Possible Solutions

Stein (2014) points out that the basis representation approach with $J \ll n$ may result in a poor likelihood approximation to the Gaussian process model for \mathbf{Z} . The problem mainly comes from the fact that using a low rank basis with $J \ll n$ may cause significant information loss regarding the smoothness of a spatial process. Due to this reason basis representation often yields a poor representation for highly

non-smooth spatial processes. Nychka et al. (2013) overcomes the limitation by using sparse covariance Σ_w defined by Markov random fields (MRF) and a sparse multiresolutional basis matrix \mathbf{K} defined by Wendland kernels. The computational efficiency due to the sparsity of both Σ_w and \mathbf{K} allows the dimensionality of \mathbf{K} to be greater than n (i.e. $J > n$) and this enables the resulting process to precisely represent highly non-smooth spatial processes. Another possible solution is to use a completely different framework that approximates the likelihood function itself instead of the process. We discuss this approach in detail in the next chapter.

2.4 Summary

In this chapter we have given a brief overview of the basics of computer model emulation and calibration using Gaussian processes. We have introduced the basic theories of Gaussian process modeling and introduced the Matérn covariance function that plays an important role in spatial process modeling. Then we have described the computer model emulation and calibration framework based on Gaussian processes and explained the computational challenges posed by high-dimensional spatial data. To motivate a possible solution to the problem we have introduced the basis representation approach and presented two approaches in detail that fall into this category: the principal components analysis and the kernel convolution. We have discussed the limitations of the basis representation approach towards the end, which motivates our investigation on the composite likelihood in the next chapter.

Composite Likelihood for Spatial Data

The composite likelihood approach approximates the log likelihood function as a sum of component log likelihoods. Each component log likelihood is a log likelihood function for a subset of data. The approach recently has gained considerable attention due to its ability to deal with complicated large data that become increasingly common in many scientific disciplines. The idea dates back to Besag (1975, 1977), which proposed to use a “pseudo-likelihood” to avoid dealing with an intractable normalizing constant in the likelihood function for the Markov random field (MRF) model. Lindsay (1988), in which the term composite likelihood is first used, generalizes the approach as a way of formulating an unbiased estimating equation. The composite likelihood framework is useful especially when (i) the original likelihood is in an intractable form, or (ii) the original likelihood is computationally expensive due to the large size or the high dimensionality of the data. Such cases often arise when the dataset of interest contains a large number of observations that are dependent on each other. Therefore modeling large spatial data is a natural area of application of the composite likelihood approach.

The key computational advantage of composite likelihood arises from using the maximum composite likelihood estimator (MCLE) in place of the MLE. Since composite likelihood is different from the original likelihood, it is a misspecified estimating equation and therefore MCLE is less efficient than the MLE. However, each component likelihood is based on the probability model corresponding to the

original likelihood and therefore the composite likelihood inherits some characteristics of the original likelihood. In particular the MCLE is a consistent estimator and, in general, asymptotically normally distributed just like the MLE. This fact makes characterizing the large sample properties of the MCLE easier than estimators based on many other estimating equations.

In this chapter the main objectives are (i) to review the various types of composite likelihoods for large spatial data, (ii) discuss the large sample properties of the MCLE, and (iii) study the use of composite likelihood for Bayesian inference. The rest of this chapter is organized as follows. We overview various types of composite likelihood for large spatial data in Section 3.1 and examine the asymptotic properties of composite likelihood estimators in Section 3.2. In Section 3.3 we then study the asymptotic properties of Bayesian estimators based on composite likelihood, and in Section 3.4 we introduce two methods to obtain MCMC samples with the correct asymptotic properties from a posterior distribution based on composite likelihood.

3.1 Overview of Composite Likelihood for Large Spatial Data

Various composite likelihood approaches have been introduced for modeling large spatial data. Depending on the form of their component likelihood these methods can be classified into two categories in general: conditional and marginal composite likelihoods (Varin, 2008).

The conditional composite likelihood approach approximates the original likelihood as a product of conditional likelihoods. Lindsay et al. (2011) discusses advantages of using conditional composite likelihood over marginal likelihood in terms of efficiency of the resulting MCLEs. Besag (1975, 1977) proposes location-wise conditional composite likelihood with log likelihood defined as

$$cl_n(\boldsymbol{\xi}) = \sum_{i=1}^n \ell(\boldsymbol{\xi}; Z(\mathbf{s}_i) | \mathcal{N}_i),$$

where $\boldsymbol{\xi}$ is a vector of all parameters, \mathcal{N}_i is the neighborhood of the location \mathbf{s}_i and

$\ell(\boldsymbol{\xi}; Z(\mathbf{s}_i)|\mathcal{N}_i)$ is the conditional likelihood for $Z(\mathbf{s}_i)$ given all observations in \mathcal{N}_i . This form is especially useful when approximating the likelihood for MRF since it removes the need to estimate the intractable constant in the original likelihood. Vecchia (1988) proposes the block-wise conditional composite likelihood such that

$$cl_n(\boldsymbol{\xi}) = \sum_{i=1}^M \ell(\boldsymbol{\xi}; \mathbf{Z}_{(i)}|\mathcal{N}_i),$$

where $\mathbf{Z}_{(i)}$ is the observations in the i th spatial block, M is the number of blocks being used, and \mathcal{N}_i is the neighborhood for the i th spatial block. See Stein et al. (2004) and Lindsay et al. (2011) for issues with choice of the neighborhood structure. The following pairwise conditional composite likelihood, introduced in a different context by Molenberghs and Verbeke (2005), can also be used to approximate the likelihood for large spatial data:

$$cl_n(\boldsymbol{\xi}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \ell(\boldsymbol{\xi}; Z_i|Z_j).$$

The marginal composite likelihood approach uses marginal likelihoods for subsets of data as its component likelihoods. Although marginal composite likelihoods tend to be less efficient than the conditional likelihood, they often provide a convenient way of approximating the original likelihood. For example, the pairwise composite likelihood defined below provides an easy way to deal with non-Gaussian data such as binary outcomes (Heagerty and Lele, 1998) or extreme values (Padoan et al., 2010) due to the intractability of the higher order joint densities for such distributions:

$$cl_n(\boldsymbol{\xi}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \ell(\boldsymbol{\xi}; Z_i, Z_j).$$

Eidsvik et al. (2013) extends it to the following pairwise block composite likelihood:

$$cl_n(\boldsymbol{\xi}) = \sum_{i=1}^{M-1} \sum_{j=i+1}^M \ell(\boldsymbol{\xi}; \mathbf{Z}_{(i)}, \mathbf{Z}_{(j)}).$$

Another example is the block composite likelihood, which is simply the sum of

marginal log likelihoods for all spatial blocks;

$$c\ell_n(\boldsymbol{\xi}) = \sum_{i=1}^M \ell(\boldsymbol{\xi}; \mathbf{Z}_{(i)}). \quad (3.1)$$

As long as the area covered by each block is large enough the block composite likelihood yields a very close approximation to the original likelihood in terms of the Kullback-Leibler divergence (Stein, 2014).

It is also possible to formulate a hybrid approach that combines the ideas of the two approaches above. For example Caragea and Smith (2006) proposes the following hybrid block composite likelihood which combines the marginal likelihood for the block means and the conditional likelihood for the blocks given their block means:

$$c\ell_n(\boldsymbol{\xi}) = \ell(\boldsymbol{\xi}; \bar{\mathbf{Z}}) + \sum_{i=1}^M \ell(\boldsymbol{\xi}; \mathbf{Z}_{(i)} | \bar{\mathbf{Z}}_{(i)}),$$

where $\bar{\mathbf{Z}}_{(i)}$ is the mean of the i th block and $\bar{\mathbf{Z}} = (\bar{\mathbf{Z}}_{(1)}, \dots, \bar{\mathbf{Z}}_{(M)})^T$ is the vector of all block means. Through a simulation study Caragea and Smith (2006) shows that this approach results in a more efficient MCLE than the block composite likelihood in (3.1).

3.2 Theoretical Issues

In this section we turn our attention to the asymptotic properties of MCLE. As we pointed out earlier the asymptotic properties of MCLE stem from those of the MLE, and therefore we first investigate the asymptotic properties of the MLE for the Gaussian process likelihood function and extend these findings to understand the large sample properties of the MCLEs.

3.2.1 Asymptotic Properties of MLE

It is well known that MLEs are consistent and asymptotically normal under fairly general regularity conditions. In this subsection we formally verify that the regularity conditions hold for the Gaussian process likelihood with Matérn covariance. The theorem below establishes two technical conditions, (i) uniform convergence of

the Hessian of the log likelihood and (ii) the asymptotic normality of the gradient of the log likelihood, which are sufficient to ensure the consistency and the asymptotic normality (Sweeting, 1980). Since these conditions are hard to directly verify, we utilize the results from Mardia and Marshall (1984) that state that these two conditions hold when the covariance function and its first and second derivatives are absolutely summable. Without a formal proof, Mardia and Marshall (1984) states that the exponential covariance function satisfies such conditions. In the theorem below we extend the statement to the Matérn covariance and provide a formal proof that verifies the absolute summability conditions for the Matérn covariance function.

Theorem 3.2.1. *Let \mathbf{Z}_n be a Matérn process on a regular lattice with increasing domain. Then $\mathbf{R}_n(\boldsymbol{\xi})^{-\frac{1}{2}}\ddot{\ell}_n(\boldsymbol{\xi})\left\{\mathbf{R}_n(\boldsymbol{\xi})^{-\frac{1}{2}}\right\}^T \xrightarrow{P} \mathbf{I}_n$ uniformly and this implies that $\frac{1}{n}\ddot{\ell}_n(\boldsymbol{\xi})$ converges to $\frac{1}{n}\mathbf{R}_n(\boldsymbol{\xi})$ uniformly in a compact ball B around $\boldsymbol{\xi}_0$ in probability, i.e.*

$$\limsup_n P\left(\sup_{\boldsymbol{\xi} \in B} \left| \frac{1}{n}\ddot{\ell}_n(\boldsymbol{\xi}) - \frac{1}{n}\mathbf{R}_n(\boldsymbol{\xi}) \right| > \epsilon\right) = 0,$$

where $\mathbf{R}_n(\boldsymbol{\xi}) = E\left(\ddot{\ell}(\boldsymbol{\xi})\right)$ and $\epsilon > 0$. This again implies $\mathbf{V}_n^{-\frac{1}{2}}\dot{c}\ell_n(\boldsymbol{\xi}) \xrightarrow{D} N(\mathbf{0}, \mathbf{I})$ for some positive definite $\mathbf{V}_n \sim \mathcal{O}(1)$. As a result, for the maximum likelihood estimator $\hat{\boldsymbol{\xi}}_n^{ML} = \arg \max_{\boldsymbol{\xi}} \ell(\boldsymbol{\xi})$,

(i) (consistency)

$$\hat{\boldsymbol{\xi}}_n^{ML} \xrightarrow{P} \boldsymbol{\xi}_0. \quad (3.2)$$

(ii) (asymptotic normality)

$$\mathbf{R}_n^{\frac{1}{2}}\left(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0\right) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}). \quad (3.3)$$

Proof. Without loss of generality, we can work on the following simplified version of the Matérn covariance function:

$$\tilde{C}_\nu(h) = \kappa h^\nu K_\nu(\phi h). \quad (3.4)$$

By Theorem 3 in Mardia and Marshall (1984), the sufficient conditions ensuring the regularity conditions in Sweeting (1980) are $\sum_{h=1}^{\infty} \tilde{C}_\nu(h) < \infty$, $\sum_{h=1}^{\infty} \frac{\partial \tilde{C}_\nu}{\partial \phi}(h) < \infty$,

$\sum_{h=1}^{\infty} \frac{\partial \tilde{C}_v}{\partial \kappa}(h) < \infty$, and $\sum_{h=1}^{\infty} \frac{\partial \tilde{C}_v}{\partial \phi \partial \kappa}(h) < \infty$.

To show that $\sum_{h=1}^{\infty} \tilde{C}_v(h) < \infty$, we first note that the modified Bessel function of the second kind can be expanded as

$$K_\nu(\phi h) = \frac{\sqrt{\pi} e^{-\phi h}}{\sqrt{2\phi h}} A_\nu(\phi h),$$

where

$$A_\nu(\phi h) = 1 + \frac{4\nu^2 - 1}{8\phi h} + \frac{(4\nu^2 - 1)(4\nu^2 - 9)}{2!(8\phi h)^2} + \frac{(4\nu^2 - 1)(4\nu^2 - 9)(4\nu^2 - 25)}{3!(8\phi h)^3} + \dots$$

Therefore the simplified Matérn covariance function in (3.4) can be rewritten as

$$\tilde{C}_v(h) = \sqrt{\frac{\pi}{2\phi}} \kappa h^{\nu - \frac{1}{2}} e^{-\phi h} A_\nu(\phi h).$$

Since $A_\nu(\phi h)$ converges to 1 as $h \rightarrow \infty$, we only need to show that $\sum_{h=1}^{\infty} h^{\nu - \frac{1}{2}} e^{-\phi h} < \infty$. The result follows by noting that

$$\lim_{h \rightarrow \infty} \frac{(h+1)^{\nu - \frac{1}{2}} e^{-\phi(h+1)}}{h^{\nu - \frac{1}{2}} e^{-\phi h}} = \lim_{h \rightarrow \infty} \left(\frac{h+1}{h} \right)^{\nu - \frac{1}{2}} e^{-\phi} < 1. \quad (3.5)$$

The result $\sum_{h=1}^{\infty} \frac{\partial \tilde{C}_v}{\partial \kappa}(h) < \infty$ immediately follows by the fact that $\tilde{C}_v(h) = \frac{\partial \tilde{C}_v}{\partial \kappa}(h)$ when $\kappa = 1$.

Next, to show $\sum_{h=1}^{\infty} \frac{\partial \tilde{C}_v}{\partial \phi}(h) < \infty$, we write the sum $\sum_{h=1}^{\infty} \frac{\partial \tilde{C}_v}{\partial \phi}(h)$ by

$$\begin{aligned} \sum_{h=1}^{\infty} \frac{\partial \tilde{C}_v}{\partial \phi}(h) &= -\frac{1}{2} \phi^{-\frac{3}{2}} \sum_{h=1}^{\infty} \tilde{C}_v(h) - \kappa \sum_{h=1}^{\infty} h^{\nu + \frac{1}{2}} e^{-2\phi h} A_\nu(\phi h) \\ &\quad + \sqrt{\frac{\pi}{2\phi}} \kappa \sum_{h=1}^{\infty} h^{\nu + \frac{1}{2}} e^{-\phi h} \frac{\partial A_\nu}{\partial \phi}(\phi h). \end{aligned}$$

We already know that the first term is finite. Showing that the second term is finite is also straightforward by using the same arguments that are used to show that $\sum_{h=1}^{\infty} \tilde{C}_v(h) < \infty$. Since $\frac{\partial A_\nu}{\partial \phi}(\phi h)$ is monotonously decreasing in h and $\sum_{h=1}^{\infty} h^{\nu + \frac{1}{2}} e^{-\phi h} < \infty$ by a similar ratio test in (3.5), the third term is also finite by Abel's convergence test. \square

3.2.2 Asymptotic Properties of MCLE

In this subsection we discuss the asymptotic properties of the MCLE. The theorem below specifies the sufficient conditions for the maximum composite likelihood estimators to be consistent and asymptotically normal. This theorem is a restatement of the general results due to White (1996) that specify the technical conditions for estimators based on estimating equations to be consistent and asymptotically normal.

Theorem 3.2.2 (White, 1996). *Let $cl_n(\boldsymbol{\xi})$ be a log composite likelihood function approximating the log likelihood function $\ell_n(\boldsymbol{\xi})$. Assume that $\frac{1}{n}\ddot{cl}_n(\boldsymbol{\xi})$ converges to $\frac{1}{n}\mathbf{Q}_n(\boldsymbol{\xi})$ uniformly in a compact ball B around $\boldsymbol{\xi}_0$ in probability, i.e.*

$$\limsup_n P \left(\sup_{\boldsymbol{\xi} \in B} \left| \frac{1}{n}\ddot{cl}_n(\boldsymbol{\xi}) - \frac{1}{n}\mathbf{Q}_n(\boldsymbol{\xi}) \right| > \epsilon \right) = 0,$$

as $n \rightarrow \infty$, where $\mathbf{Q}_n(\boldsymbol{\xi}) = E \left(\ddot{cl}(\boldsymbol{\xi}) \right)$ and $\epsilon > 0$, and $\mathbf{V}_n^{\frac{1}{2}} \dot{cl}_n(\boldsymbol{\xi}) \xrightarrow{D} N(\mathbf{0}, \mathbf{I})$ as $n \rightarrow \infty$ for some symmetric positive definite matrix $\mathbf{V}_n \sim \mathcal{O}(1)$. Then for the maximum composite likelihood estimator $\hat{\boldsymbol{\xi}}_n^{CL} = \arg \max_{\boldsymbol{\xi}} cl_n(\boldsymbol{\xi})$,

(i) (consistency)

$$\hat{\boldsymbol{\xi}}_n^{CL} \xrightarrow{P} \boldsymbol{\xi}_0. \quad (3.6)$$

(ii) (asymptotic normality)

$$\mathbf{G}_n^{\frac{1}{2}}(\boldsymbol{\xi}_0) \left(\hat{\boldsymbol{\xi}}_n^{CL} - \boldsymbol{\xi}_0 \right) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}). \quad (3.7)$$

where $\mathbf{G}_n(\boldsymbol{\xi}) = \mathbf{Q}_n^{-1} \mathbf{P}_n \mathbf{Q}_n^{-1}(\boldsymbol{\xi})$ is the Godambe information matrix, $\mathbf{P}_n(\boldsymbol{\xi}) = \text{Cov} \left(\dot{cl}(\boldsymbol{\xi}) \right)$, and $\mathbf{Q}_n(\boldsymbol{\xi}) = E \left(\ddot{cl}(\boldsymbol{\xi}) \right)$.

Proof. The proof of this theorem can be found in White (1996) (see Theorem 3.4 and 6.4). \square

Note that the asymptotic distribution in 3.2.2 (ii) specifies the limiting distribution of MCLE under the original probability model, not the probability model corresponding to the composite likelihood. Under the original model the second Bartlett identity fails due to the misspecification of the likelihood function and

$\mathbf{G}_n^{-1}(\boldsymbol{\xi}_0)$ does not reduce to $\mathbf{Q}_n^{-1}(\boldsymbol{\xi}_0)$. One can measure the information loss caused by using the composite likelihood instead of the original likelihood by comparing their asymptotic covariances $\mathbf{R}_n^{-1}(\boldsymbol{\xi}_0)$ and $\mathbf{G}_n^{-1}(\boldsymbol{\xi}_0)$ since they are computed under the same probability model.

Verifying the technical conditions poses non-trivial challenges. However, as long as the component likelihoods are well defined to capture the important characteristics of the data, the corresponding maximum composite likelihood estimators (MCLEs) are (weakly) consistent and asymptotically normal under the same regularity conditions for the MLEs (Lindsay, 1988). This property stems from the fact that each component likelihood is a valid likelihood for a part of the data or some aggregated data. For this reason, in most cases, verifying the regularity conditions for the original likelihood is enough to establish the consistency and the asymptotic normality of the MCLEs. As we have already seen in the previous section the regularity condition holds for Gaussian processes with a Matérn covariance function. Therefore, one can expect that the estimators that maximize the composite likelihoods introduced in Section 3.1 have the asymptotic properties described in 3.2.2. See Caragea and Smith (2006), Eidsvik et al. (2013) for simulation study results verifying this argument. A more formal theoretical treatment of this result is a potential area for future research.

3.3 Bayesian Inference Using Composite Likelihood

In this section we discuss use of the composite likelihood methods in Bayesian inference. The application to composite likelihood to Bayesian inference has been limited relative to its application in maximum likelihood contexts but there are a few instances, notably Ribatet et al. (2012) and Shaby (2014). For the same motivations described in Section 3.1 in some cases it is useful to replace the original likelihood with a composite likelihood to formulate the posterior density. In this thesis we call the resulting posterior density a composite posterior. The same natural question then arises as in the case of MCLE: What is lost by using the composite posterior in place of the original posterior and, if any, how do we take

the information loss into account in our inference? To answer this question, we first investigate the asymptotic properties of the composite posterior function itself and the corresponding composite posterior estimators. The former will tell us what is the property of MCMC sample from the composite posterior (“what we will get from MCMC”), and the latter will tell us what is the property of the posterior estimator under the original probability model (“what we should have gotten from MCMC”). Then the rest of this section will describe how we use those asymptotic results to formulate a valid statistical inference procedure.

3.3.1 Asymptotic Properties of Composite Posterior Estimators

Under the essentially same conditions as those in Theorem 3.2.2 the posterior estimators have similar asymptotic properties to those of the MCLEs. Building upon the results from Amemiya (1985) and White (1996), Chernozhukov and Hong (2003) provides a set of technical conditions that leads to consistency and asymptotic normality. We list a modified conditions that imply those conditions which are slightly more restrictive but much easier to verify in practice, especially when working on composite likelihood for spatial data.

Assumption 3.3.1. *For a composite likelihood $cl_n(\boldsymbol{\xi})$, the following regularity conditions hold:*

(i) *(Identifiability, Chernozhukov and Hong, 2003) For any positive δ we can find $\epsilon > 0$ such that*

$$\liminf_{n \rightarrow \infty} P \left(\sup_{\|\boldsymbol{\xi} - \boldsymbol{\xi}_0\| \geq \delta} \frac{1}{n} (cl_n(\boldsymbol{\xi}) - cl_n(\boldsymbol{\xi}_0)) \leq -\epsilon \right) = 1$$

(ii) *(Compactness) $\boldsymbol{\xi} \in \Xi$ where Ξ is a convex compact subset of R^d .*

(iii) *(Differentiability and continuity) $cl_n(\boldsymbol{\xi})$ is twice continuously differentiable. $\mathbf{P}_n(\boldsymbol{\xi})$ and $\mathbf{Q}_n(\boldsymbol{\xi})$ are positive definite.*

(iv) *(Normality of the first derivative) $\mathbf{V}_n^{\frac{1}{2}}(\boldsymbol{\xi}_0) \dot{cl}_n(\boldsymbol{\xi}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{I})$ for some $\mathbf{V}_n(\boldsymbol{\xi}_0)$.*

(v) (Uniform convergence of the information matrix)

$$\limsup_n P \left(\sup_{\boldsymbol{\xi} \in B} \left| \frac{1}{n} \ddot{c} \ell_n(\boldsymbol{\xi}) - \frac{1}{n} \mathbf{Q}_n(\boldsymbol{\xi}) \right| > \epsilon \right) = 0.$$

Under these assumptions the composite posterior density converges in a similar manner as the original posterior does in the Bernstein-von Mises theorem. To state this result we first need to define the total variation norm for density functions below.

Definition 3.3.1. For a probability density function f of random vector \mathbf{x} and a positive constant $\alpha \geq 0$, the total variation is given by

$$\|f\|_{TV} = \int (1 + \|\mathbf{x}\|^\alpha) |f(\mathbf{x})| d\mathbf{x}.$$

Now we are ready to state the consistency results for the composite posterior estimator.

Theorem 3.3.1. With a prior density that is continuous and positive at $\boldsymbol{\xi}_0$ the composite posterior density $\pi(\boldsymbol{\xi}|\mathbf{Z}_n)$ based on $c \ell_n(\boldsymbol{\xi})$ satisfies

$$\|\pi(\boldsymbol{\xi}|\mathbf{Z}_n) - \pi^*(\boldsymbol{\xi})\|_{TV(\alpha)} \xrightarrow{P} 0,$$

for any $\alpha \geq 0$, where $\pi^*(\boldsymbol{\xi})$ is the normal density with mean $\boldsymbol{\xi}_0 + \mathbf{Q}_n^{-1}(\boldsymbol{\xi}_0) \dot{c} \ell(\boldsymbol{\xi}_0)$ and covariance $\mathbf{Q}_n^{-1}(\boldsymbol{\xi}_0)$.

Proof. Assumptions 2.4.1-2.4.5 implies the Assumptions 1, 3, and 4 in Chernozhukov and Hong (2003). The results follow by Theorem 1 in Chernozhukov and Hong (2003). \square

The consistency theorem above shows two important characteristics of the composite posterior estimator: First, the composite posterior estimator converges to the true value in the sense that its posterior density accumulates around the true value as the sample size grows into infinity, due to the fact that $\mathbf{Q}_n^{-1}(\boldsymbol{\xi}_0) \rightarrow \mathbf{0}$. Second, this is probably more relevant to our interest here, for a fairly large n , the posterior sample for the parameter $\boldsymbol{\xi}$ follows a normal distribution with mean $\boldsymbol{\xi}_0 + \mathbf{Q}_n^{-1}(\boldsymbol{\xi}_0) \dot{c} \ell(\boldsymbol{\xi}_0)$ and covariance $\mathbf{Q}_n^{-1}(\boldsymbol{\xi}_0)$.

Now we turn our attention to the asymptotic normality of the composite posterior estimator. The theorem below tells us the asymptotic distribution of the composite posterior estimators, such as the mean, median, and mode.

Theorem 3.3.2. *Let $\hat{\boldsymbol{\xi}}^B = \arg \max_{\boldsymbol{\xi}} \pi_n(\boldsymbol{\xi})$ be the posterior mode of $\pi_n^B(\boldsymbol{\xi})$ (the same result holds for the posterior mean or median), then*

$$\mathbf{G}_n^{\frac{1}{2}}(\boldsymbol{\xi}_0) \left(\hat{\boldsymbol{\xi}}_n^B - \boldsymbol{\xi}_0 \right) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}),$$

as $n \rightarrow \infty$

Proof. Assumptions 2.4.1-2.4.5 imply Assumptions 1, 2 and 4 in Chernozhukov and Hong (2003), and the loss function that results in the posterior mode satisfies Assumption 3. The results follow by Theorem 2 in Chernozhukov and Hong (2003). \square

This result is about the asymptotic distribution of the posterior estimator under the original probability model as in Theorem 3.2.2. The asymptotic covariance of the posterior estimator is the same as that of the MCLE, and this indicates that the composite likelihood dominates the composite posterior as the sample size tends to infinity.

One difficulty in applying the results above is to verify the regularity conditions in 3.3.1, especially (iv) and (v). In general, those conditions are satisfied when the MCLE for the composite likelihood used in defining the composite posterior is consistent and asymptotically normal (Shaby, 2014). The following lemma partially explains why this statement generally holds.

Lemma 3.3.1. *Let $c\ell_n(\boldsymbol{\xi})$ be a composite likelihood function which is twice continuously differentiable. Assume that (3.6) and (3.7) in Theorem 3.2.2 and Assumption 3.3.1 (v) hold, then the condition (iv) in 3.3.1 holds.*

Proof. Since $c\ell_n(\boldsymbol{\xi})$ is twice continuously differentiable, its first derivative $\dot{c}\ell_n(\boldsymbol{\xi})$ at the true value $\boldsymbol{\xi}_0$ can be rewritten as

$$\begin{aligned} \dot{c}\ell_n(\boldsymbol{\xi}_0) &= \dot{c}\ell_n(\hat{\boldsymbol{\xi}}_n) + \ddot{c}\ell_n(\boldsymbol{\xi}_0)(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0) + o_p(1) \\ &= \mathbf{Q}_n(\boldsymbol{\xi}_0)(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0) + o_p(1), \end{aligned}$$

because $\dot{c}\ell_n(\hat{\boldsymbol{\xi}}_n) = \mathbf{0}$ and $\mathbf{Q}_n^{-1}(\boldsymbol{\xi}_0)\ddot{c}\ell_n(\boldsymbol{\xi}_0) \xrightarrow{P} \mathbf{I}_d$ uniformly. Since $\mathbf{Q}_n(\boldsymbol{\xi}_0)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)$ is asymptotically normally distributed, the result follows. \square

3.4 Posterior Sample Adjustment

An important observation on the results in the Theorems 3.3.1 and 3.3.2 is that the posterior estimate based on MCMC sample generated from the composite posterior does not have the correct asymptotic distribution. The composite posterior density converges to the density with the covariance of $\mathbf{Q}_n^{-1}(\boldsymbol{\xi}_0)$ while the asymptotic covariance of the composite posterior estimator under the original probability model is $\mathbf{G}_n^{-1}(\boldsymbol{\xi}_0)$. As a result, the MCMC sample based on the composite posterior incorrectly represent the uncertainty associated with the parameter estimation.

Shaby (2014) and Ribatet et al. (2012) propose similar approaches to adjusting the composite posterior density to account for this discrepancy. The common idea in both approaches is to adjust the posterior sample $\boldsymbol{\xi}$ for each iteration of the MCMC chain using the following formula:

$$\tilde{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_n^B + \mathbf{D}(\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_n^B),$$

with some constant $d \times d$ matrix \mathbf{D} . Shaby (2014) proposes to use $\mathbf{D} = \mathbf{Q}_n^{-1}\mathbf{P}_n^{\frac{1}{2}}\mathbf{Q}_n^{\frac{1}{2}}$ while Ribatet et al. (2012) suggests to use $\mathbf{D} = \mathbf{Q}_n^{\frac{1}{2}}(\mathbf{Q}_n\mathbf{P}_n^{-1}\mathbf{Q}_n)^{\frac{1}{2}}$. Shaby (2014) verifies that those two approaches yield essentially the same result.

3.5 Discussion

In this chapter we have reviewed the composite likelihood approaches for modeling high-dimensional spatial data and study the asymptotic properties of the estimators based on the composite likelihoods. We have also studied the asymptotic properties of Bayesian estimators based on composite posteriors and have described how to obtain MCMC sample with the correct variation using those properties.

So far we have been focusing on reviewing two approaches to modeling high-dimensional spatial data: the basis representation approaches and the composite

likelihood approach. In the next two chapters (4 and 5), we introduce new climate model calibration methods based on these two approaches and apply them to real climate model calibration problems. In Chapter 4, in addition to the methodological development, we demonstrate that using high-dimensional spatial data provides a better calibration result than using aggregated data and therefore show that developing calibration methods using high-dimensional data can contribute to reducing uncertainties about projections of future climate.

Chapter 4

Fast Dimension-Reduced Climate Model Calibration and the Effect of Data Aggregation*

How will the climate system respond to anthropogenic forcings? One approach to this question relies on climate model projections. Current climate projections are considerably uncertain. Characterizing and, if possible, reducing this uncertainty is an area of ongoing research. We consider the problem of making projections of the North Atlantic meridional overturning circulation (AMOC). Uncertainties about climate model parameters play a key role in uncertainties in AMOC projections. When the observational data and the climate model output are high-dimensional spatial data sets, the data are typically aggregated due to computational constraints. The effects of aggregation are unclear because statistically rigorous approaches for model parameter inference have been infeasible for high-resolution data. Here we develop a flexible and computationally efficient approach using principal components and basis expansions to study the effect of spatial data aggregation on parametric and projection uncertainties. Our Bayesian reduced-dimensional calibration approach allows us to study the effect of complicated error structures and data-model discrepancies on our ability to learn about climate model parameters from high-dimensional data. Considering high-dimensional spatial ob-

*This chapter is based on a manuscript published in the *Annals of Applied Statistics*, authored by Won Chang, Murali Haran, Roman Olson, and Klaus Keller.

servations reduces the effect of deep uncertainty associated with prior specifications for the data-model discrepancy. Also, using the unaggregated data results in sharper projections based on our climate model. Our computationally efficient approach may be widely applicable to a variety of high-dimensional computer model calibration problems.

4.1 Introduction

Computer models play an important role in understanding complex physical processes in modern science and engineering. They are particularly important in climate science where computer models, complex deterministic systems used to model the Earth system, are used both to study climate phenomena as well as make projections about the future. A major source of uncertainty in climate projections is due to uncertainties about model parameters. Parameter calibration involves characterizing our knowledge about a model parameter by using observational data. Here we use calibration to refer to a statistical method that summarizes information about a parameter in terms of a probability distribution. In this distribution parameter values that generate output more compatible with observational data are assigned higher probabilities than parameters less compatible with observations. Calibration of the parameters using observational data is hence one avenue to reduce the uncertainty in future projections. A number of issues and challenges arise when performing statistical calibration of model parameters. Because each run of the computer model is computationally expensive, computer model output is typically obtained for a relatively small sample of parameter values. Furthermore, the model output at each parameter setting may be high-dimensional and in the form of spatial fields. A sound statistical approach to this problem needs to simultaneously address the spatial dependence in the data and model outputs, account for various sources of uncertainty, and remain computationally efficient. Computational efficiency is key in order to utilize all the relevant observations at the appropriate scale; previous methods for climate model calibration have relied on heavy data aggregation, thereby potentially discarding valuable information.

The scientific problem motivating our statistical analysis is the projection of the future state of the North Atlantic meridional overturning circulation (AMOC)

in response to anthropogenic climate change. The AMOC is a large-scale ocean circulation that transports cold and dense water equatorward in the deep North Atlantic, and warm and salty water poleward in the upper layers of the North Atlantic. The AMOC might show a persistent weakening in response to anthropogenic forcing. Because the AMOC plays an important role in heat and carbon transport, an AMOC weakening is projected to have considerable impacts on climate, and, in response, on natural and human systems (cf. Alley et al., 2007; Keller et al., 2005, 2007). We use previously published perturbed physics ensemble runs (Sriver et al., 2012) of the University of Victoria Earth System Climate Model (UVic ESCM) (Weaver et al., 2001) to set up the calibration problem. Specifically, the runs model transient behavior of the climate system over the years 1800-2100. Each run starts from a control climate, obtained by running the model from the same initial condition to equilibrium at preindustrial conditions. Vertical ocean mixing is important in projecting the AMOC (Wunsch and Ferrari, 2004), but most of the mixing occurs on scales below that of the UVic ESCM, hence mixing is “parameterized” (cf. Weaver et al., 2001; Schmittner et al., 2009; Goes et al., 2010) using a “vertical background diffusivity” (K_{bg}). The AMOC projections depend on the K_{bg} parameter values (e.g. Goes et al., 2010). The value of K_{bg} is uncertain; it therefore needs to be calibrated using observations of the climate that are informative about K_{bg} (cf. Goes et al., 2010; Bhat et al., 2012).

Here, we calibrate K_{bg} using observations of ocean potential temperature from the World Ocean Atlas 2009 (Antonov et al., 2010; Locarnini et al., 2010). The World Ocean Atlas is a gridded data product generated by interpolation of instrumental observations. The observational data at irregularly distributed locations are interpolated onto a regular grid by Barnes interpolation (Barnes, 1964). The parameter K_{bg} affects the depth of the oceanic pycnocline (Gnanadesikan, 1999) and the AMOC (Bryan, 1987; Goes et al., 2010). As a consequence, models with different K_{bg} values are expected to result in different ocean temperature distributions. Ocean temperatures are therefore informative about K_{bg} . Note that neither data assimilation nor calibration play any role in producing the World Ocean Atlas data set and the data product does not depend on any assumptions about vertical diffusivity.

Both the UVic ESCM output and the observational data are spatial data sets

with more than 60,000 spatial locations. Of particular interest is how data aggregation affects the calibration result for K_{bg} . Often observations of climate and climate model outputs are 3-D spatial fields. When the spatial data sets are large it is common practice to aggregate them into 1-D or 2-D patterns (Sansó and Forest, 2009; Drignei et al., 2008; Forest et al., 2008; Goes et al., 2010; Bhat et al., 2012; Olson et al., 2012; Schmittner et al., 2009) either to avoid computational issues or because the skill of the models at higher resolution may not always be trusted.

An important and interesting question is what information, if any, is lost by this data aggregation. Aggregating data for model calibration may increase or decrease the model parameter uncertainty depending on the relative importance of several processes. Data aggregation may lead to information loss which would result in larger uncertainties about the calibrated parameters. On the other hand, data aggregation can potentially reduce the magnitude of model errors, for instance due to unresolved variability or structural errors in the climate models, which could in turn lead to smaller uncertainties about the calibrated parameters. We hypothesize for the scientific questions and models we consider here that data aggregation may lead to considerable loss of information resulting in increased uncertainties about model parameters. Increased uncertainties about parameters propagates to increased uncertainty in climate projections, which can impact risk- and decision-analysis.

We adopt a Gaussian process based approach to the calibration problem (Sacks et al., 1989; Kennedy and O’Hagan, 2001). Gaussian processes provide flexible statistical interpolators or “emulators” of the computer model across various parameter settings and are therefore attractive for climate model calibration (cf. Sansó and Forest, 2009; Bhat et al., 2012). Unfortunately, the likelihood evaluations involved in fitting such models can become prohibitive with high-dimensional spatial data due to the expensive matrix operations involved. Current approaches for high-dimensional computer model calibration can reduce the computational burden and make likelihood evaluation feasible for moderately large datasets (spatial fields observed at a few thousand locations) or datasets that are on a regular and complete grid (Higdon et al., 2008; Bhat et al., 2012; Bayarri et al., 2007). However, to our knowledge, no current calibration approach can overcome the computational challenge of dealing with large spatial data sets (more than tens of thousands of

data points) on an incomplete grid. Here an incomplete grid refers to one with a large number of missing points (about 40%).

The impact of data aggregation on climate model calibration is a largely unanswered question due to the inability of existing methods to analyze large spatial datasets of both computer model output and observations. Throughout this manuscript we will use “large” to refer to data sets that comprise over tens of thousands of spatial observations. Here we develop a computationally efficient approach that handles large data sets. This approach gives us the freedom to carry out a careful study of the effects of data aggregation, for example comparing calibration based on unaggregated three-dimensional data with calibration based on aggregated two-dimensional or one-dimensional data. Our approach also enables one to investigate the interaction between data aggregation and data-model discrepancies and errors when inferring computer model parameters. In our simulated examples, we have shown that the method can handle complicated model-observation discrepancy processes without sacrificing computational efficiency.

The remainder of this chapter is organized as follows. In Section 4.2 we provide a description of the data set. In Section 4.3 we describe our two-stage framework for climate model calibration and the associated computational challenges. In Section 4.4 we propose a general model calibration approach in a reduced dimensional space that uses a combination of principal components and a basis representation to overcome computational challenges. In Section 4.5 we provide implementation details and in Section 4.6 we discuss the results from simulated examples and real data. We conclude this chapter with caveats and future directions in Section 4.7.

4.2 Data Description

Our goal is to build an emulator based on spatial output from UVic ESCM and to calibrate vertical ocean diffusivity (K_{bg}) using ocean potential temperature data. The UVic ESCM runs are 3-dimensional patterns of the mean ocean potential temperature over 1955–2006 at 250 parameter settings. The parameters controlling model outputs are vertical ocean diffusivity (K_{bg}), anthropogenic aerosol scaling factor (A_{scl}), and climate sensitivity (C_s). Note that we converted longwave radiation feedback factor, which is one of the original input parameters for UVic, into

C_s using a simple spline fit. We refer to Srivier et al. (2012) for the design points and details of the ensemble runs.

To avoid problems related to model artifacts and sparse sampling, we excluded data beyond 60°N and 80°S and 3000m in depth (Key et al., 2004; Schmittner et al., 2009; Bhat et al., 2012). UVic ESCM outputs are on a 77 (latitude) \times 100 (longitude) \times 13 (depth) grid, but the number of locations that have non-missing observations is 65,595. The missing values occur because there is no ocean at the locations in the UVic ESCM representation. At each grid point we compute a temporal mean over the time period of 1955–2006 to average out the effect of unresolved internal variability.

The observational data are on a 180 (latitude) \times 360 (longitude) \times 33 (depth) grid, and we remap this observed data into the UVic model grid using a linear interpolation using only the nearby points. See Figure A.1 and A.2 in the Appendix A for comparison between the UVic model grid and the observational data grid. This results in a relatively small reduction to 61,051 data points. The model output locations are also adjusted accordingly. We convert the observed in-situ temperature field into the potential temperature field in order to (i) have the same measurement unit with UVic ESCM output and (ii) adjust the effect of pressure on ocean temperature. We obtain potential temperature from the in-situ temperature (Locarnini et al., 2010) and salinity fields (Antonov et al., 2010) using the UNESCO equation of state (UNESCO, 1981) following Bryden (1973) and Fofonoff (1977). During the conversion procedure, we assume a simplified ocean pressure field varying as a function of latitude and depth (Lovett, 1978). As we do for the UVic ESCM output, we compute a temporal mean over 1955–2006 at each location. Figure 4.1 shows examples of UVic ESCM model runs and the converted observational data.

4.3 Model Calibration Framework

Our computer model calibration framework consists of two stages, (i) model emulation and (ii) parameter calibration (Bayarri et al., 2007; Bhat et al., 2012). First, we construct an emulator that interpolates computer model outputs at different parameter settings using Gaussian random fields (Sacks et al., 1989). This

can be viewed as statistical interpolation or “kriging” (Cressie, 1994) in the computer model parameter space. Second, we infer the computer model parameters by relating observational data to computer model output using the emulator, while considering observational error and allowing for systematic discrepancies between the model and observations (Kennedy and O’Hagan, 2001). Note that this two-stage approach has some advantages over fully Bayesian methods that combine the two stages into a single inferential step. By constructing an emulator solely based on computer model output (Rougier, 2008; Bhat et al., 2012; Liu et al., 2009), this two-stage approach ensures that inference in the emulation stage is not contaminated by model discrepancy and observational error. In addition, separating the emulation stage from calibration provides an easier way to diagnose the accuracy of an emulator. Furthermore, computations are faster and parameter identifiability problems are reduced.

Let $Y(\mathbf{s}, \boldsymbol{\theta})$ denote the computer model output at the spatial location $\mathbf{s} = (\text{longitude}, \text{latitude}, \text{depth}) \in \mathcal{S} \subseteq \mathbb{R}^3$ and the model parameter setting $\boldsymbol{\theta} \in \Theta$, where \mathcal{S} is the spatial domain of the process and Θ is the computer model parameter space, typically a subset of a unidimensional or multidimensional Euclidean space. In our calibration problem, $\Theta \subset \mathbb{R}^3$ since there are three input parameters. Furthermore, $Z(\mathbf{s})$ is the corresponding observation at the spatial location \mathbf{s} . Since each run of the climate model is computationally expensive, we can obtain computer model outputs only for a relatively small number of design points p . We denote these design points in the parameter space by $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p \in \Theta$. Let $\mathbf{Y}_i \in \mathbb{R}^n$ be the computer model output at each parameter setting $\boldsymbol{\theta}_i$ for $i = 1, \dots, p$. Each computer model output $\mathbf{Y}_i = (Y(\mathbf{s}_1, \boldsymbol{\theta}_i), \dots, Y(\mathbf{s}_n, \boldsymbol{\theta}_i))^T$ is a spatial process observed at n different spatial locations $(\mathbf{s}_1, \dots, \mathbf{s}_n)$. In our calibration problem, $n = 61,501$ and $p = 250$. Let \mathbf{Y} be the vector of concatenated computer model outputs such that $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_p)^T$. We denote the observed spatial process at n locations by $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$. Note that we assume that the locations for each model output and observation data are the same. If they are different, one can interpolate either of them depending on which one has a higher resolution. Our objective is to infer the parameter $\boldsymbol{\theta}$ by combining information from \mathbf{Z} and \mathbf{Y} .

4.3.1 Two-Stage Emulation and Calibration

We first outline our general framework for emulation and calibration.

Model Emulation Using Gaussian Random Fields

As described in Bhat et al. (2012), a standard approach to approximate the climate model output is using a Gaussian process such that

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

with a $np \times b$ covariate matrix \mathbf{X} and a vector of regression coefficient $\boldsymbol{\beta}$. The covariates in \mathbf{X} are the spatial locations (e.g. latitude, longitude, and depth) and the climate parameters. The covariate matrix \mathbf{X} contains all the spatial coordinates and the parameter settings used to define the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\xi}_y)$. The vector $\boldsymbol{\xi}_y$ contains all the parameters determining the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\xi}_y)$. In our application, since the mean term $\mathbf{X}\boldsymbol{\beta}$ is set to $\mathbf{0}$ (see Section 4.4.2.1 for more details), we fit a Gaussian random field to \mathbf{Y} by finding the maximum likelihood estimate (MLE) of only $\boldsymbol{\xi}_y$, denoted by $\hat{\boldsymbol{\xi}}_y$.

The fitted Gaussian random field defines the probability model for the computer model output at any location $\mathbf{s} \in \mathcal{S}$ and parameter setting $\boldsymbol{\theta} \in \Theta$. Therefore, the Gaussian process model provides a predictive distribution of computer model output at any untried value of $\boldsymbol{\theta}$ given the existing output \mathbf{Y} (Sacks et al., 1989). We denote the resulting interpolated process by $\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y})$ and call it an emulator process. This approach automatically provides a quantification of interpolation uncertainty.

Model Calibration Using Gaussian Random Field Model

Once an emulator $\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y})$ is available, we model the observational data \mathbf{Z} ,

$$\mathbf{Z} = \boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}) + \boldsymbol{\delta} + \boldsymbol{\epsilon}, \quad (4.1)$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ is independently and identically distributed observational error and $\boldsymbol{\delta}$ is a data-model discrepancy term. The discrepancy $\boldsymbol{\delta}$ is also modeled as a Gaussian process, thus $\boldsymbol{\delta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_d(\boldsymbol{\xi}_d))$ with a spatial covariance matrix

$\Sigma_d(\boldsymbol{\xi}_d)$ between the locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ and a vector of covariance parameters $\boldsymbol{\xi}_d$. The details regarding the specification of the covariance function are provided in Section 4.4.2.2. This discrepancy term is crucial for parameter calibration (cf. Bayarri et al., 2007; Bhat et al., 2010). Note that this problem is ill-posed without any prior information for $\boldsymbol{\xi}_d$, so an informative prior is necessary. Our inference for $\boldsymbol{\theta}$, $\boldsymbol{\xi}_d$ and σ^2 is based on their resulting posterior distribution.

4.3.2 Challenges with High-Dimensional Data

High-dimensional datasets pose considerable computational challenges due to the expensive likelihood function calculations that involve high-dimensional matrix computations. For instance, in the calibration problem described in Section 4.5, the dimensionality of the model output and the observational data is $n = 984$ in the 2-D case and $n = 61,051$ in the 3-D case, with $p = 250$. The latter example involves prohibitive computations with naïve implementations (discussed and explained in Section A.1 in the Appendix A. For instance, with n -dimensional climate model outputs at p different parameter settings, evaluation of the likelihood function requires $\mathcal{O}(n^3 p^3)$ operations. Therefore, numerical methods such as Newton-Raphson or MCMC algorithms become infeasible.

4.4 Model Calibration with High-Dimensional Spatial Data

We develop a dimension reduction approach based on spatial basis functions to increase computational efficiency. Spatial basis functions can map high-dimensional data into a low-dimensional space (Bayarri et al., 2007) and find a representation of the probability model that results in lower computational cost for likelihood evaluations (Higdon et al., 2008; Bhat et al., 2012). Since there may be a trade-off between parsimony and accurate inference, it is crucial to find a set of spatial basis functions that gives a computationally feasible likelihood formulation without considerable loss of information. Below, we review drawbacks to the current approaches in the context of high-dimensional spatial data and propose a new approach to overcome these limitations.

4.4.1 Current Approaches

Various methods have been introduced to overcome computational challenges with models for high-dimensional spatial data. These methods may be roughly grouped into the following three categories: low-rank representations, likelihood approximations, and sparse covariance approximations. Low-rank representation methods such as kernel convolution (Higdon, 1998), Gaussian predictive process (Banerjee et al., 2008), and fixed rank kriging (Cressie and Johannesson, 2008) approximate spatial processes using a set of basis functions and typically reduce the computational costs by using algorithms for patterned covariance matrices, for instance the Sherman-Morrison-Woodbury formula. Likelihood approximation methods substitute the expensive likelihood function with a relatively inexpensive approximation such as the Whittle likelihood (Fuentes, 2007), or composite likelihood (Vecchia, 1988; Stein et al., 2004; Caragea and Smith, 2006; Eidsvik et al., 2013). Sparse covariance approximation methods such as covariance tapering (Furrer et al., 2006), Gaussian Markov random field approximations (Lindgren et al., 2011; Simpson et al., 2012), and lattice kriging (Nychka et al., 2013) introduce sparsity in the covariance or precision matrix thereby allowing for fast computations using sparse matrix algorithms.

A few different approaches to climate model calibration with multivariate computer model outputs have been developed in recent years, including Bayarri et al. (2007), Higdon et al. (2008), and Bhat et al. (2012). These approaches, however, are not readily applicable to the 3-D model output and observations we consider here due to the following reasons. First, in spite of the gains in computational efficiency likelihood evaluations remain computationally prohibitive. The computational cost of a single likelihood evaluation in the emulation step in Bhat et al. (2012) scales as $\mathcal{O}(nJ^2)$ where J is the number of knots for the kernel basis. In Higdon et al. (2008), the computational cost scales with $\mathcal{O}(p^3 J_y^3)$ where J_y is the number of principal components used to represent the data. For the 3-D calibration problem we consider, n is 61,051 and J should be more than $3(K_v) \times 3(A_{scl}) \times 3(C_s) \times 3(\text{depth}) \times 20(\text{longitude}) \times 15(\text{latitude}) = 24,300$ to ensure the number of knots to be at least three and greater than 20% of the design points for each dimension. The number of principal components J_y needs to be at least 20 to have more than 90% of explained variance. Second, the transformation

based on the basis matrix may not be applicable to two or three-dimensional spatial data. Using a wavelet transformation (Bayarri et al., 2007) requires the same dyadic (a power of 2) number of data points for each spatial dimension, and the data need to be on a regular grid without missing values; irregular data and missing values are common in climate model calibration problems. In addition, conducting Bayesian inference on the joint posterior distribution may pose difficulties, both computationally as well as in terms of prior specification and identifiability issues. For example, Higdon et al. (2008) requires estimating $4 \times J_y + 1$ parameters, which translates to an 81-dimensional distribution for the 3-D case in Section 4.5.

4.4.2 Reduced Dimensional Model Calibration

Our method to overcome the aforementioned challenges relies on (i) representing the spatial field using a principal component basis, and (ii) emulating each principal component separately. Instead of using a principal component basis to reduce the complexity of matrix computation as in Higdon et al. (2008), we use it to map the computer model outputs into a low-dimensional space and construct an emulator in that space directly. Since the principal components are uncorrelated by construction, we can build the emulator by constructing a 1-dimensional Gaussian process for each principal component in parallel. Fitting Gaussian random fields for each principal component requires estimation of only five parameters (see below). The likelihood evaluations involve covariance matrices of size $p \times p$. These features allow us to construct the emulator in a computationally efficient and highly automated manner. Moreover, since the principal component transformation can be applied to non-dyadic spatial data with irregular locations, it has a broader range of application than wavelet transformations. In the calibration step, we develop an approach to map the observational data into a low dimensional space.

4.4.2.1 Computer Model Emulation

The first step of this approach is to find the basis matrix for computer model output. We consider the computer model outputs as an n -dimensional dataset with p replications and find the principal component basis. Let \mathbf{M} denote the

$p \times n$ matrix storing the computer model outputs $\mathbf{Y}_1, \dots, \mathbf{Y}_p$ such that

$$\mathbf{M} = \begin{pmatrix} \mathbf{Y}_1^T \\ \vdots \\ \mathbf{Y}_p^T \end{pmatrix}. \quad (4.2)$$

Following the standard process of finding principal components, we first preprocess the computer model outputs to make the column means of the matrix \mathbf{M} all 0's. Applying singular value decomposition (SVD), we find the scaled eigenvectors $\mathbf{k}_1 = \sqrt{\lambda_1} \mathbf{e}_1, \dots, \mathbf{k}_p = \sqrt{\lambda_p} \mathbf{e}_p$ where $\lambda_1 > \lambda_2 > \dots > \lambda_p$ are the ordered eigenvalues and $\mathbf{e}_1, \dots, \mathbf{e}_p$ are the corresponding eigenvectors of the covariance matrix of \mathbf{M} , where $J_y \ll p$ is the number of principal components that we decide to use in the emulator. One can choose the number of principal components by looking at the proportion of explained variation given by $\frac{\sum_{i=1}^{J_y} \lambda_i}{\sum_{i=1}^p \lambda_i}$. We define the basis matrix for computer model output by $\mathbf{K}_y = (\mathbf{k}_1, \dots, \mathbf{k}_{J_y})$.

For each parameter setting $\boldsymbol{\theta}_i$ ($i = 1, \dots, p$), the first J_y principal components $\mathbf{Y}_i^R = (Y_{i1}^R, \dots, Y_{iJ_y}^R)^T$ are computed as

$$\mathbf{Y}_i^R = (\mathbf{K}_y^T \mathbf{K}_y)^{-1} \mathbf{K}_y^T \mathbf{Y}_i.$$

Let $\mathbf{Y}^R = (\mathbf{Y}_1^R, \dots, \mathbf{Y}_p^R)^T$, hence each element of this matrix $\{\mathbf{Y}^R\}_{ij} = Y_{ij}^R$ is the j th principal component at the i th computer model parameter setting. Since the columns in \mathbf{K}_y are orthogonal, the principal components found here are uncorrelated to each other and this leads us to a parallelized emulator construction that is similar to the wavelet transformation approach in Bayarri et al. (2007). For each j th principal component across the parameter settings (i.e. for each j th column of \mathbf{Y}^R), we construct a Gaussian random field with the squared exponential covariance function such that

$$\text{Cov}(Y_{kj}^R, Y_{lj}^R) = \kappa_{y,j} \exp \left(- \sum_{i=1}^3 \frac{|\theta_{ik} - \theta_{il}|^2}{\phi_{y,ij}^2} \right) + \zeta_{y,j} 1(\boldsymbol{\theta}_k = \boldsymbol{\theta}_l), \quad (4.3)$$

with partial sill $\kappa_{y,j}$, nugget $\zeta_{y,j}$, and range parameters $\boldsymbol{\phi}_{y,j} = (\phi_{y,1j}, \phi_{y,2j}, \phi_{y,3j})^T$. Leave-10-percent-out cross-validation experiments with 50 different randomly generated parameter configuration indicate that the squared exponential covariance

shows a better fit than alternatives such as the exponential covariance (Figure A.3 in the Appendix A). Note that this choice of covariance function may yield a fitted parameter field that is too regular, resulting in imprecise estimation of the range parameters $\phi_{y,ij}$, $i = 1, \dots, 3$, $j = 1, \dots, J_y$. However, our purpose here is not to estimate the range parameters precisely; the range parameters depend on the distance in the parameter space, which is a somewhat arbitrary notion. Note also that the mean term of each Gaussian process used here is set to be zero, since each of the principal components has zero mean across the parameter settings.

We denote the collection of emulator parameters for the j th principal component by $\boldsymbol{\xi}_{y,j} = (\kappa_{y,j}, \zeta_{y,j}, \boldsymbol{\phi}_{y,j})^T$. One can construct the emulator by finding the MLE $\hat{\boldsymbol{\xi}}_{y,j}$ for each j separately. The emulator $\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}^R)$ is the collection of predictive processes of J_y principal components at $\boldsymbol{\theta}$ defined by the covariance function (4.3) and the MLEs $\hat{\boldsymbol{\xi}}_{y,1}, \dots, \hat{\boldsymbol{\xi}}_{y,J_y}$. Note that even though we construct the emulator in terms of the principal components, we can make a projection \mathbf{Y}^* in the original space at a new parameter setting $\boldsymbol{\theta}^*$ by computing

$$\mathbf{Y}^* = \mathbf{K}_y \boldsymbol{\eta}(\boldsymbol{\theta}^*, \mathbf{Y}^R).$$

To summarize, the emulation step uses the data $\mathbf{Y}_1^R, \dots, \mathbf{Y}_{J_y}^R$ of dimension p and compute MLEs $\hat{\boldsymbol{\xi}}_{y,1}, \dots, \hat{\boldsymbol{\xi}}_{y,J_y}$. Hence, the computational cost is reduced from $\mathcal{O}(n^3 p^3)$ to $\mathcal{O}(J_y p^3)$ when compared to a naïve approach. The resulting fitted model is then used for the calibration step as described in the following section.

4.4.2.2 Computer Model Calibration

Using $\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}^R)$, the emulator for the principal components, we reformulate the model for observational data in (4.1) as

$$\mathbf{Z} = \mathbf{K}_d \boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}^R) + \mathbf{K}_d \boldsymbol{\nu} + \boldsymbol{\epsilon},$$

where $\mathbf{K}_d \boldsymbol{\nu}$ is a kernel convolution representation (Higdon, 1998) of the discrepancy $\boldsymbol{\delta}$. $\boldsymbol{\nu}$ is a vector of independent and identically distributed Normal random variates at $J_d \ll n$ locations, $\boldsymbol{\nu} \sim N(\mathbf{0}, \kappa_d \mathbf{I}_{J_d})$. $\mathbf{a}_1, \dots, \mathbf{a}_{J_d} \in \mathcal{S}$. The variance parameter κ_d determines the magnitude of discrepancy, and the range parameters $\phi_{d,1}, \phi_{d,2} > 0$

specify the bandwidth of kernels. We define the kernel basis by

$$(\mathbf{K}_d)_{ij} = \exp\left(-\frac{g(s_{1i}, s_{2i}, a_{1j}, a_{2j})}{\phi_{d,1}} - \frac{|s_{3i} - a_{3j}|}{\phi_{d,2}}\right), \quad (4.4)$$

where s_{ki} and a_{kj} are the k th elements of \mathbf{s}_i and \mathbf{a}_j respectively. Our choice for the knot points $\mathbf{a}_1, \dots, \mathbf{a}_{J_d}$ are on a grid of 15.6 degrees in latitude, 36 degrees in longitude, and 429m in depth. The design of these points does not affect the resulting process unless chosen to be too sparse. The geodesic distance function measures the great circle distance between two points on the Earth's surface. The function $g(s_{1i}, s_{2i}, a_{1j}, a_{2j})$ is given by

$$r \arccos(\sin(s_{2i}) \sin(a_{2j}) + \cos(s_{2i}) \cos(a_{2j}) \cos |s_{1i} - a_{1j}|),$$

where r is the radius of Earth (6378km). By following Higdon et al. (2008), the range parameters are pre-specified by scientific expert judgment; this reduces computations and identifiability issues (See Section 4.5 for specification of these parameters). The kernel function in (4.4) yields a valid covariance under geodesic distance since it is strictly positive definite on a sphere (Gneiting, 2013). We assumed separability for distance along the surface and distance along the depth. The resulting process is approximately twice differentiable (Zhu and Wu, 2010), which produces a reasonable model for discrepancy. Even though the discrepancy model implies an isotropic discrepancy process (Higdon, 2002), the resulting process is flexible enough to capture the general trend in the discrepancy.

Instead of using the model (4.4) directly, we conduct calibration with reduced-dimensional data for computational efficiency. Let \mathbf{Z}^R be a reduced version of the original data such that

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

where $\mathbf{K} = (\mathbf{K}_y \mathbf{K}_d)$. The probability model of \mathbf{Z}^R is

$$\mathbf{Z}^R \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_\eta \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_\eta & \mathbf{0} \\ \mathbf{0} & \kappa_d \mathbf{I}_{J_d} \end{pmatrix} + \sigma^2 (\mathbf{K}^T \mathbf{K})^{-1}\right), \quad (4.5)$$

where $\boldsymbol{\mu}_\eta$ and $\boldsymbol{\Sigma}_\eta$ are the mean and covariance given by the emulator $\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}^R)$. It is often helpful to apply singular value decomposition to \mathbf{K}_d and use the first $J_d^{PC} \ll J_d$ eigenvectors \mathbf{K}_d^{PC} in place of \mathbf{K}_d to find \mathbf{Z}^R . In addition to the obvious computational advantage, this often results in better inference since it corresponds to a regularized estimate given by ridge regression (see Hastie et al., 2009, pg. 66); this was corroborated by our extensive simulation studies.

Note that the term $\sigma^2(\mathbf{K}^T\mathbf{K})^{-1}$ in (4.5) automatically adjusts the contribution of each principal component to the calibration result. This can be illustrated by considering the model without the discrepancy, and the variance in the likelihood function is simply $\boldsymbol{\Sigma}_\eta + \sigma^2(\mathbf{K}_y^T\mathbf{K}_y)^{-1}$. Since $(\mathbf{K}_y^T\mathbf{K}_y)^{-1}$ is a diagonal matrix and its i th diagonal element is the reciprocal of the i th eigenvalue, $(\mathbf{K}_y^T\mathbf{K}_y)^{-1}$ inflates the variance of principal components with small eigenvalues. Therefore, the principal components with smaller explained variance will have less effect on the calibration result.

We now briefly examine the covariance structure implied by our model. Using the leading J_y principal components, the covariance between computer model outputs at two different spatial and parametric coordinates $(\mathbf{s}_1, \boldsymbol{\theta}_1)$ and $(\mathbf{s}_2, \boldsymbol{\theta}_2)$ can be written as

$$\begin{aligned} \text{Cov}(Y(\mathbf{s}_1, \boldsymbol{\theta}_1), Y(\mathbf{s}_2, \boldsymbol{\theta}_2)) &\approx \text{Cov}\left(\sum_{i=1}^{J_y} \sqrt{\lambda_i} \mathbf{e}_i(\mathbf{s}_1) w_i(\boldsymbol{\theta}_1), \sum_{j=1}^{J_y} \sqrt{\lambda_j} \mathbf{e}_j(\mathbf{s}_2) w_j(\boldsymbol{\theta}_2)\right) \\ &= \sum_{i=1}^{J_y} \lambda_i \mathbf{e}_i(\mathbf{s}_1) \mathbf{e}_i(\mathbf{s}_2) \text{Cov}(w_i(\boldsymbol{\theta}_1), w_i(\boldsymbol{\theta}_2)), \end{aligned}$$

where $\mathbf{e}_i(\cdot)$ is the i th eigenfunction satisfying

$$\int \text{Cov}(Y(\boldsymbol{\theta}_1, \mathbf{s}_1), Y(\boldsymbol{\theta}_2, \mathbf{s}_2)) \mathbf{e}_i(\mathbf{s}_2) d\mathbf{s}_1 = \lambda_i \mathbf{e}_i(\mathbf{s}_2) \text{Cov}(w_i(\boldsymbol{\theta}_1), w_i(\boldsymbol{\theta}_2)),$$

with the corresponding eigenvalue λ_i . We let $w_i(\cdot)$ denote the Gaussian process of i th principal component with the covariance function defined in (4.3). The leading eigenfunctions give the best approximation among all possible orthogonal bases since it minimizes the total mean square error (Jordan, 1961). Since we can assume different covariance functions for each principal component process, our

model can yield a non-separable space-parameter covariance function. In contrast, if we were to assume separability such that

$$\text{Cov}(Y(\boldsymbol{\theta}_1, \mathbf{s}_1), Y(\boldsymbol{\theta}_2, \mathbf{s}_2)) = C_s(\mathbf{s}_1, \mathbf{s}_2)C_\theta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2),$$

for some positive definite covariance functions C_s and C_θ , the covariance function for the i th principal component process becomes

$$\text{Cov}(w_i(\boldsymbol{\theta}_1), w_i(\boldsymbol{\theta}_2)) = C_\theta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\lambda_i.$$

The detailed derivation is provided in Section A.2 in the Appendix A. The separability assumption therefore results in a restrictive covariance structure such that the correlation functions for all principal component processes are the same. Hence even though our reduced dimensional approach utilizes a covariance that is easy to specify, it provides a richer class of covariance functions than a separable covariance structure. Our cross-validation studies show that our assumption is adequate for emulating the computer model (see Section 4.5 for details).

Priors

We estimate the joint density of $\boldsymbol{\theta}$, κ_d , and σ^2 using the Metropolis-Hastings algorithm. Following Bayarri et al. (2007), we allow for additional flexibility by estimating the partial sill parameters $\kappa_{y,1}, \dots, \kappa_{y,J_y}$ for the emulator. Prior specification for the parameters in the observational model is straightforward. The discrepancy variance κ_d and the observational error variance σ^2 receive inverse-gamma priors with small shape parameter values. The prior for each parameter is a uniform distribution over a broad range or determined by scientific knowledge. In order to stabilize the inference, we put an informative prior to encourage $\kappa_{y,1}, \dots, \kappa_{y,J_y}$ to vary around their estimated values in the emulation stage. See Section 4.5 for more details about prior specifications for our problem.

Computing

The computational costs may be summarized as follows:

1. Find the basis matrix $\mathbf{K}_y = (\sqrt{\lambda_1}\mathbf{e}_1, \dots, \sqrt{\lambda_{J_y}}\mathbf{e}_{J_y})$ by computing the singular value decomposition of \mathbf{M} in (4.2). This computation is of order $\mathcal{O}(n^3)$, but needs to be done only once.
2. Compute \mathbf{Y}_R where its i th row is the transpose of $(\mathbf{K}_y^T \mathbf{K}_y)^{-1} \mathbf{K}_y^T \mathbf{Y}_i$.
3. Construct a Gaussian random field for each column of \mathbf{Y}_R by finding the MLE $\hat{\boldsymbol{\xi}}_{y,i}$ for each $i = 1, \dots, J_y$. The computational cost is of order $\mathcal{O}(J_y p^3)$ for each likelihood evaluation.
4. Compute $\mathbf{Z}_R = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T \mathbf{Z}$. The computational complexity of this step is $\mathcal{O}((J_y + J_d)^3)$.
5. Using Metropolis-Hastings, draw an MCMC sample of $\boldsymbol{\theta}$, σ^2 , κ_d and $\kappa_{y,1}, \dots, \kappa_{y,J_y}$ from the joint posterior distribution based on the model in (4.5). The computational cost for each likelihood evaluation is of order $\mathcal{O}((J_y + J_d)^3)$.

The overall cost of our implementation is $\mathcal{O}(pJ_y^3)$ for the emulation step and $\mathcal{O}((J_y + J_d)^3)$ for the calibration step.

4.5 Implementation Details

We apply our method to data at three different aggregation levels. In the 1-D case, we compute the vertical means at $n = 13$ different depth points (Goes et al., 2010). In the 2-D case, the zonal means are computed at $n = 984$ latitude and depth points (Bhat et al., 2012). We use the original pattern without any aggregation in the 3-D case ($n = 61,051$). The number of principal components is determined to have more than 90% of the explained variation. The number of components is 5 for the 1-D, 10 for the 2-D, and 20 for the 3-D case. We also tried using 10 principal components for the 1-D, 20 for the 2-D, and 30 for the 3-D case to have more than 95% explained variation, but did not find any improvement in the calibration result. These results coincide with our exploratory analysis using scree plots. For example, the scree plot for the 3-D case shown in Figure A.4 in the Appendix indicates that the non-leading principal components beyond the

first 10 components have extremely small eigenvalues comparing to the first few components. Similar results hold for the 2-D and the 1-D cases.

We use all 250 design points in the parameter space to build the emulator. We conducted leave-10-percent out cross validation and the results show that our emulator can predict the model output precisely. For the 2-D (latitude-depth) case, for instance, the emulator based on principal components can reproduce well the spatial pattern at any given parameter setting. More specifically, we randomly held out 25 model runs from the model output and then predicted these “hold outs” based on the remaining 225 model runs using our emulator. An example of our results is shown in Figure 4.6, which indicates that the predicted output and the original output are essentially indistinguishable. Other cross validation results, including for the 3-D case, are similar. In addition, we found that the root mean squared error was very small relative to the scale of the data. We also tested the prediction performance of our emulator for the principal components using uncorrelated standardized prediction errors (Bastos and O’Hagan, 2009) (see Section A.3 in the Appendix A for details). The graphical diagnostics in Figure A.5 in the Appendix A show that our emulators predict the processes of leading principal components reasonably well.

We fix C_s and A_{scl} at the default values of the UVic ESCM in the calibration stage and made an inference only for K_{bg} . The default values are 1 for A_{scl} and 3.819 for C_s . One may choose to integrate out these two parameters, but since the ocean temperature field lacks strong information about A_{scl} and C_s , their estimated posterior densities are overly dispersed. This introduces unnecessary bias in the estimate of K_{bg} due to the highly non-linear relationship between climate parameters (Olson et al., 2012), thus we decided not to integrate out those two parameters.

Following Bhat et al. (2012), we assume a flat prior with a broad range for K_{bg} , from 0.05 to 0.55. The variance for the observational error (σ^2) and the model discrepancy (κ_d) receive inverse-Gamma priors, and we denote them by $\text{IG}(a_\nu, b_\nu)$ and $\text{IG}(a_z, b_z)$. We set the shape parameters for them to be $a_\nu = 2$ and $a_z = 2$. To check the sensitivity of our approach to prior specifications, we tried four different combinations, (2, 2), (2, 100), (100, 2), and (100, 100) for b_ν and b_z . The emulator variances $\kappa_{y,1}, \dots, \kappa_{y,J_y}$ also receive inverse-Gamma priors with a shape parameter

of 5. The scale parameters are determined to have modes at their estimated values in the emulation stage. Because of parameter identifiability problems which in turn affected computation, we fixed the range parameter for depth at 3000m and for surface as 4800km. The knot points $(\mathbf{a}_1, \dots, \mathbf{a}_{J_d})$ are on a grid of 15.6 degrees in latitude, 36 degrees in longitude, and 429m in depth. The design of these points does not affect the resulting process unless chosen to be too sparse. We found that a wide range of the different settings for these parameters gave the same inference result for K_{bg} and hence our particular choices did not affect the results. The number of knot locations for the discrepancy kernel is 800 in the 3-D case, 80 in the 2-D case, and 13 in the 1-D case. The number of principal components used for the discrepancy is 200 in the 3-D case, 20 in the 2-D case, and 5 in the 1-D case. The number of principal components was determined using standard practice – by ensuring that at least 95% of the variability in the data was explained in each case. In order to check the robustness of our results, we tried different numbers of principal components. For example, when we increased the number of principal components to 300 in the 3-D case, 30 in the 2-D case and 8 in the 1-D case we found that we obtained virtually identical calibration results.

Finally, we note that we ran an MCMC algorithm with 25,000 iterations for the calibration step. We carefully checked our results by comparing summaries (e.g. posterior density estimates) based on the first 15,000 runs with those obtained from the entire 25,000 runs and verified that our MCMC-based estimates are reliable.

4.6 Results

4.6.1 Computational Benefit

The biggest challenge in the considered analysis is the computational cost of evaluating the likelihood function in the 3-D case, which requires dealing with 61,051-dimensional spatial datasets. To our knowledge, current approaches cannot address this problem with reasonable computational effort (discussed below). In the emulation stage, the required number of principal components is about 20 for the 3-D case for reasonable accuracy and this means we still need to invert a $(p \times J_y) \times (p \times J_y) = 5,000 \times 5,000$ covariance matrix for the likelihood evaluation

in the emulation stage if one uses the formulation due to Higdon et al. (2008). Moreover, the number of parameters to be estimated is $20 \times 4 + 1 = 81$ ($J_y \times$ the number of parameters for each Gaussian process emulator + one nugget parameter), and this also possibly increases computational cost significantly. The method of Bhat et al. (2012) requires multiplication of a $J \times n = 24,300 \times 61,501$ matrix into another $n \times J = 61,501 \times 24,300$ matrix in the likelihood evaluation, and this makes the likelihood evaluation computationally prohibitive.

The approximate computational time in emulation stage for the 1-D, 2-D and 3-D cases for different methods are illustrated in Figure 4.2. Our approximation was derived as follows:

$$(\text{computing time for our approach}) \times \frac{(\text{complexity for a method})}{(\text{complexity for our approach})}.$$

The computing time for our approach is for the PORT routine in R (Gay, 1983; David, 1990) run on a system with Intel Xeon E5450 Quad-Core 3.0 GHz (without parallelization). Note that this approximated computing time for other approaches are optimistic; we believe that they will probably take longer than indicated. We use this to suggest that even when viewing the other approaches' cost optimistically, our method provides dramatically reduced computational time. We also note here that the computing time for our approach can be further reduced by parallelization. We describe some experimental results on computing time reduction using parallel computing in Section A.4 in the Appendix A.

4.6.2 The Effect of Data Aggregation on Climate Model Calibration

In order to study the effect of data aggregation on climate model calibration, we conducted a study with pseudo-observational data. The simulated data are generated as follows:

1. Choose the 3-D pattern of UVic ESCM output with $K_{bg} = 0.2$, $A_{scl} = 1.5$, and $C_s = 3.976$ as the synthetic truth. The values for A_{scl} and C_s were selected based on the fact that they were the closest parameter values to the default values for the UVic model; the value for K_{bg} was obtained from the

posterior mode from previously published work (cf. Bhat et al., 2012).

2. Compute three different 3-D patterns of residuals between the observational data (Wold Ocean Atlas 2009 data described above) and the UVic model outputs with $K_{bg} = 0.1$, $K_{bg} = 0.2$, and $K_{bg} = 0.3$. The values for A_{scl} and C_s are the same as in Step 1. Average them over each location to compute a pseudo-residual. This is a more realistic and challenging residual than one obtained by simulation from a simple error model, for example a realization from a Gaussian process model. For brevity, we describe here just this most challenging case; our methods worked even better in terms of posterior variance when error processes were assumed to be simpler.
3. Superimpose the pseudo residual on the synthetic truth to construct pseudo observational data in 3-D.
4. Aggregate the 3-D pseudo observational data into 2-D and 1-D by integrating the ocean temperature with respect to water volume.

The calibration results based on this simulated example are shown in Figure 4.3. The sensitivity test indicates that the posterior distribution of K_{bg} in the 1-D and the 2-D cases rely on the specification of priors. This deep uncertainty (Groves and Lempert, 2007) is drastically reduced when the full data set (3-D) is used. A comparison result based on the real data from Ocean Atlas 2009 is shown in Figure 4.4. As in the simulated example, the calibration results based on the 3-D data are more robust to the prior specification.

In addition to the experiment above we also examined the effect of random sampling of spatial locations, which may be considered a reasonable alternative to averaging over particular dimensions. As an illustrative example we randomly chose 1,300 grid points from the 3-D locations using simple random sampling and calibrated K_{bg} based on only the selected points. By repeating the same experiment with 10 different random samples we found that calibration results are substantially different across the samples (see Figure A.6 in the Appendix A). These results indicate that sampling of spatial locations may introduce additional sampling errors to the calibration results; using all available data points is therefore desirable.

Using the full pattern of the 3-D data has important benefits as it drastically reduces the deep uncertainty due to different prior specifications. We hypothesize that this is because the full non-aggregated spatial patterns contain more information about both the observational error and the discrepancy. In order to reflect the uncertainty due to prior choice to our density estimate for K_{bg} , we show in Figure 4.5 the posterior distributions when the prior is assumed to be with equal probability any one of the 4 priors considered, along with the resulting AMOC projections. We define AMOC projection as the annual maximum value of the meridional overturning streamfunction in the Atlantic between 0° and 70°N . The corresponding projections for AMOC change between 1970 to 1999 mean and 2070 to 2099 mean indicate that the unaggregated pattern gives a much narrower 95% predictive interval than the aggregated ones. Therefore, data aggregation increases the deep uncertainty surrounding AMOC projection, and using unaggregated data reduces uncertainty regarding the future behavior of the AMOC.

4.7 Discussion

4.7.1 Caveats

A general issue with principal components is also worth considering in this context: the principal components for the computer model outputs are selected based on explained variation, and thus there is no guarantee that these leading principal components carry the most important information about the climate parameters. However, our extensive study of the effect of changing the number of principal components suggests that this is not problematic in our context. Our results are consistent with the recent theoretical results in Artemiou and Li (2009) that suggest that there is a low probability that other (non-leading) principal components will have a strong correlation with the climate parameters. We hypothesize that our principal components based approach does not lose valuable information about the climate parameters.

In fact the first principal component shown in Figure A.7 captures some important characteristics of the model runs related to K_{bg} and AMOC. The latitude-depth pattern for the first principal component seems to represent the differences in

vertical diffusion between ensemble members; in simple upwelling diffusion models under low diffusion more warmth is concentrated near the surface at low latitudes (see e.g. Appendix B in Kriegler, 2005). The latitude-longitude pattern is related to difference between model runs in (i) the effect of AMOC in the North Atlantic, which is expected to cause warmer temperatures in the North Atlantic near the coast of Greenland at higher K_{bg} , and (ii) other large scale circulations such as Western Boundary Currents (primarily Kuroshio) and Eastern Boundary Currents (e.g., Peru current).

In the discrepancy model, one important simplifying assumption is the separability between surface and depth effects. Our simulated example shows that the separability assumption provides a good approximation to the realistic discrepancy process. Non-separable covariance function that combines geodesic distance and Euclidean distance remains as an avenue of ongoing and alive research and subject of future work. Furthermore, our study of calibration with simulated examples shows that even though the number of K_{bg} settings at which the model is run is relatively sparse, there is enough information to reliably calibrate K_{bg} based on our emulator.

Our study is also subject to the usual caveats with respect to scientific conclusions. First, we ignore the interpolation uncertainty when we compute the density of AMOC projection based on the density of K_{bg} . Second, the result is based on a single data set, and thus we cannot fully evaluate the effect of structural uncertainty due to the model-observation discrepancy and unresolved natural variability cannot be accounted for; this variability could impact conclusions as well (Olson et al., 2013). These caveats, of course, apply to almost all existing approaches to climate model calibration and projection.

4.7.2 Summary

We have considered here the problem of calibrating a climate model parameter, K_{bg} , in an Earth System Model of Intermediate complexity by using spatial observations of the potential temperature. In order to study the effects on both calibration and climate projections of using unaggregated versions of the data, we develop new methods that are computationally tractable for calibration with

high-dimensional spatial data sets. Using our methods, we show that utilizing 3-D spatial data reduces the uncertainty about K_{bg} and is more robust to various prior specifications than calibration based on 2-D or 1-D aggregated versions of the data. The results suggest that using unaggregated data is valuable for reducing deep uncertainty associated with different priors. We note that we have tested our method in several other real data calibration problems and have obtained similar results. For example, we carried out calibration for K_{bg} using a completely different set of observations, CFC-11 (Bhat et al., 2012), and found again that using the 3D pattern sharpened our inference about K_{bg} when compared to 2D patterns.

We have demonstrated here that our computer model calibration approach is computationally efficient even when dealing with high-dimensional data. By exploiting the orthogonality of a principal components decomposition of the data, this method can keep the computational cost affordable for high-dimensional data with more than 60,000 spatial locations and 250 parameter settings. In addition, our simulated examples show that our approach can handle complicated model-observation discrepancies. The method can be easily extended to allow for calibration with multiple tracers – we can simply consider the variance-covariance matrix for all tracers and use its principal components to build an emulator. For the 2-D case, the posterior densities from using a separable covariance structure are not very different from our PCA-based approach; in fact, the bias appears to be slightly smaller for our approach (Figure A.8 in the Appendix A). Moreover, our method results in a sharper density, which is an important criterion for calibration performance (cf. Gneiting et al., 2007). For larger data, it is more challenging to devise a simple but flexible approach that scales as well as the PCA-based method. We note here that our approach can be applied to even larger datasets. In climate science data sets consisting of millions of data points are common. Our calibration approach in principle applies immediately to such data. A potential computational bottleneck is the SVD. However, we note that the SVD needs to be performed only once and we sidestep issues related to memory management since we do not have to store the large matrices. Furthermore, high-dimensional SVD is an active area of research and hence computationally efficient approaches are being developed (cf. Halko et al., 2011; Baglama and Reichel, 2005).

In the context of computer model calibration for making climate projections in

general, we find that spatial data aggregation appears to generally lead to larger uncertainties (Figure 4.5), which is consistent with what we had hypothesized about the model and observations. This implies that the effect of information loss on the calibrated parameter is more important than the effect of reduction in model errors by aggregation. From a climate projections and decision-making perspective, the projections we obtain can be used as an input to integrated assessment or economic models; reduced uncertainties may therefore have tangible implications. Our method is immediately available for use with other climate models and observational products. By virtue of unlocking the full wealth of previously untapped information in large three-dimensional datasets, our method has a strong potential to improve projections of a host of policy-relevant climate variables.

In this chapter we have developed a reduced dimension approach based on basis representation described in Chapter 2. An alternative approach to efficient computation for computer model emulation and calibration with high-dimensional spatial data is composite likelihood (described for spatial data in Chapter 3). Composite likelihood-based computer model emulation and calibration is the subject of the next chapter.

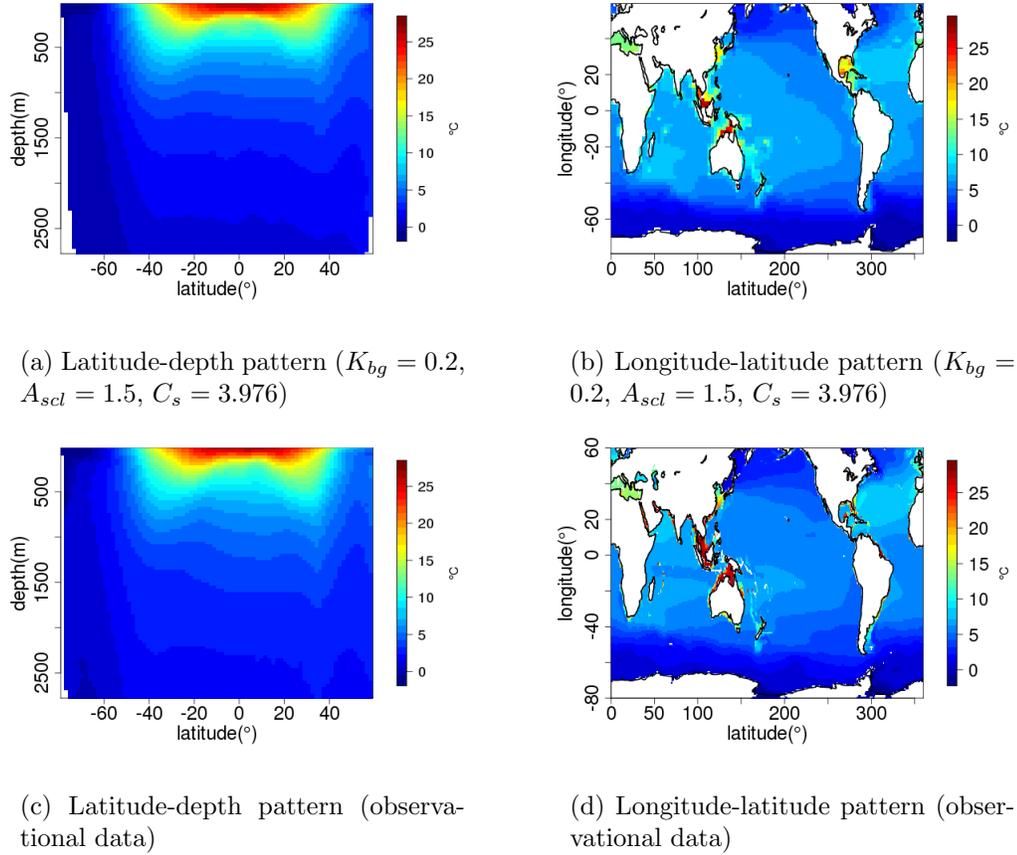


Figure 4.1: Plots of ocean temperature patterns averaged over 1955–2006. A UVic ESCM run (Srifer et al., 2012, the first row) and the observational data from World Ocean Atlas 2009 (Antonov et al., 2010; Locarnini et al., 2010, the second row). The left column shows the latitude-depth profiles of zonal mean, and the right column displays the longitude-latitude profiles of vertical mean. Note that the model run shown here is an example of 250 model runs described in Section 4.2.

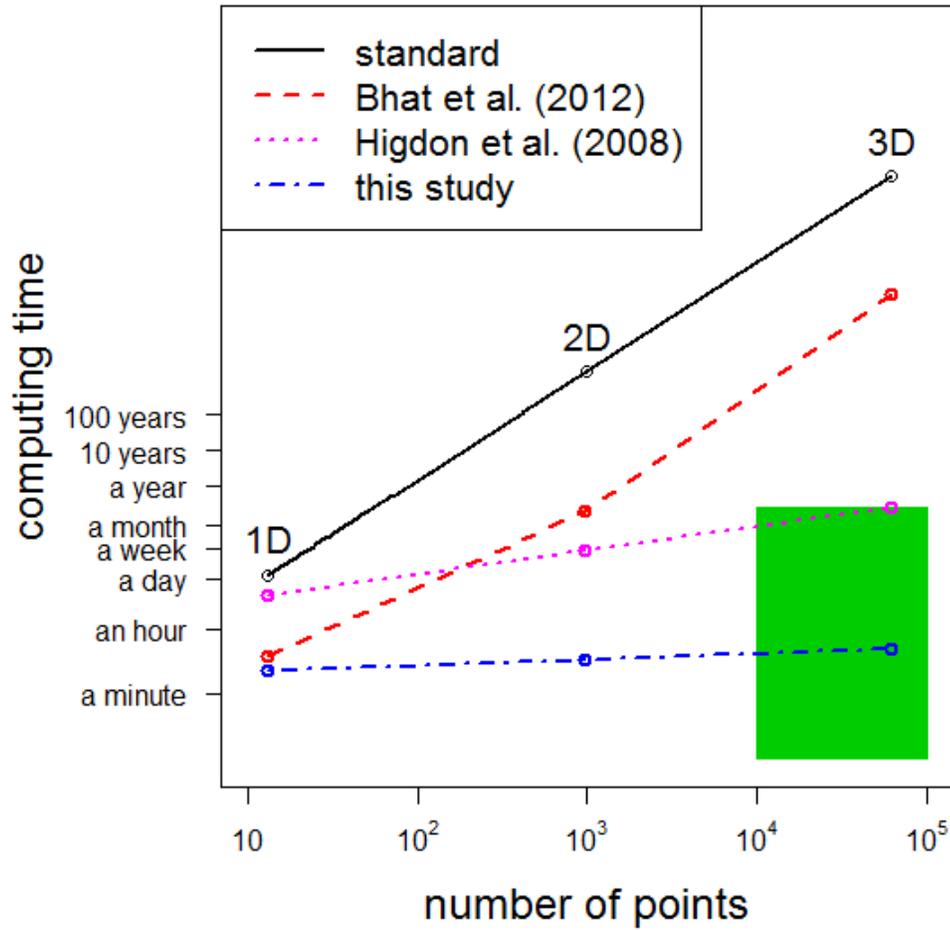


Figure 4.2: Comparison of computational costs for the emulation step between the current approaches and the new approach. The green box near the bottom right corner shows computing times that are practical, ranging from one second to three months.

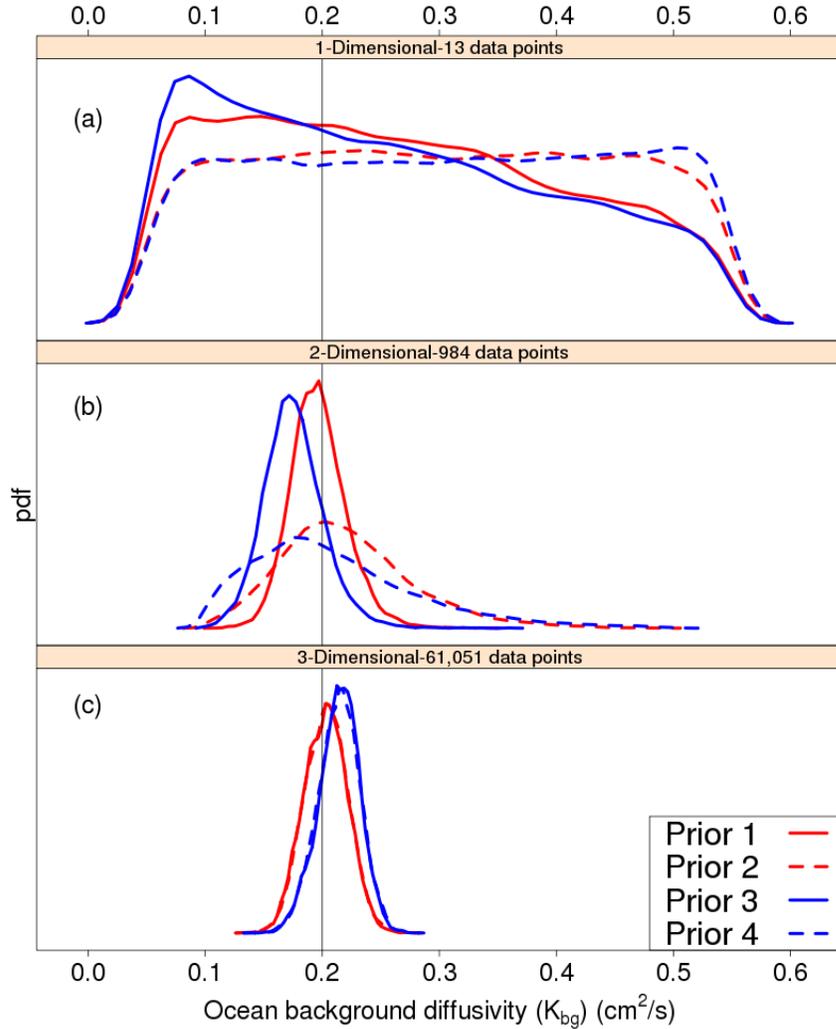


Figure 4.3: Prior sensitivity test in the simulated example. Calibration of K_{bg} value based on (a) 1-D depth profile, (b) 2-D latitude-depth pattern (c) 3-D non-aggregated data. Each line represents posterior density from four different priors: ($b_v = 2, b_z = 2$) (Prior 1, solid red line), ($b_v = 2, b_z = 100$) (Prior 2, dashed red line), ($b_v = 100, b_z = 2$) (Prior 3, solid blue line), and ($b_v = 100, b_z = 100$) (Prior 4, dashed blue line). The solid vertical line represents the true value of K_{bg} in the synthetic truth. b_v and b_z are hyperparameters of κ_d and σ^2 respectively in (4.5).

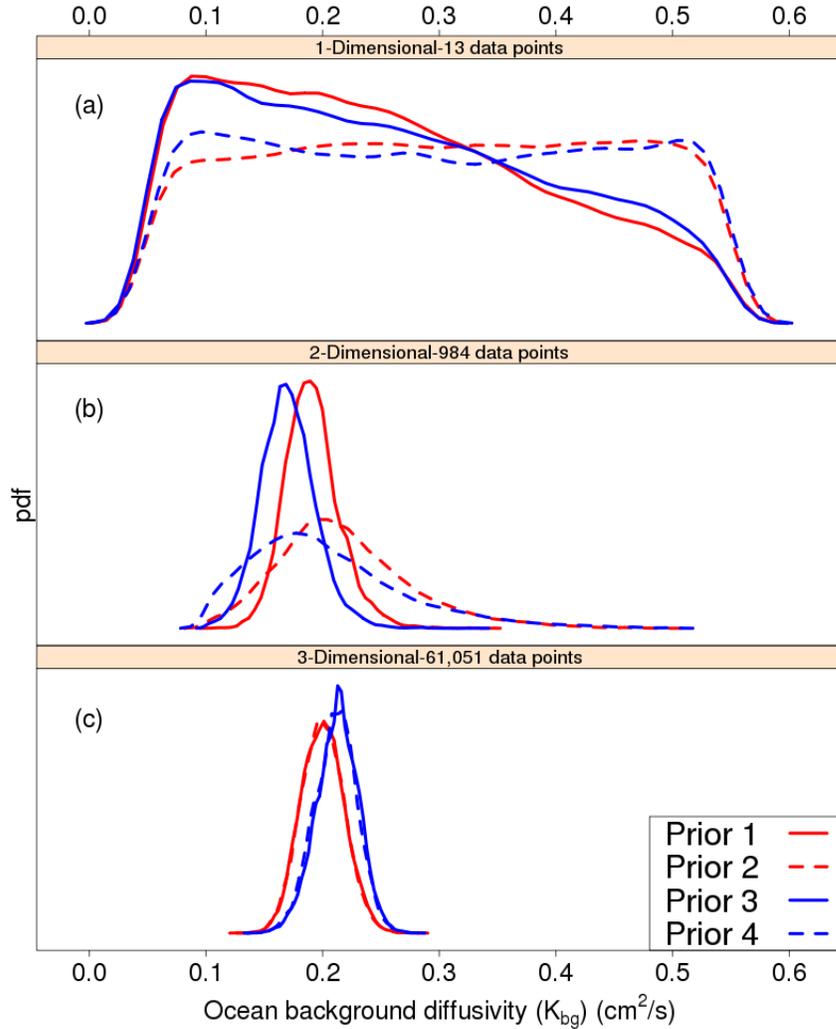


Figure 4.4: Prior sensitivity test using observational data from the World Ocean Atlas 2009 (Antonov et al., 2010; Locarnini et al., 2010). Calibration results based on (a) 1-D depth profile, (b) 2-D latitude-depth pattern, (c) 3-D non-aggregated data. Each line represents posterior density from four different priors: $(b_v = 2, b_z = 2)$ (Prior 1, solid red line), $(b_v = 2, b_z = 100)$ (Prior 2, dashed red line), $(b_v = 100, b_z = 2)$ (Prior 3, solid blue line), and $(b_v = 100, b_z = 100)$ (Prior 4, dashed blue line), where b_v and b_z are hyperparameters of κ_d and σ^2 respectively in (4.5).

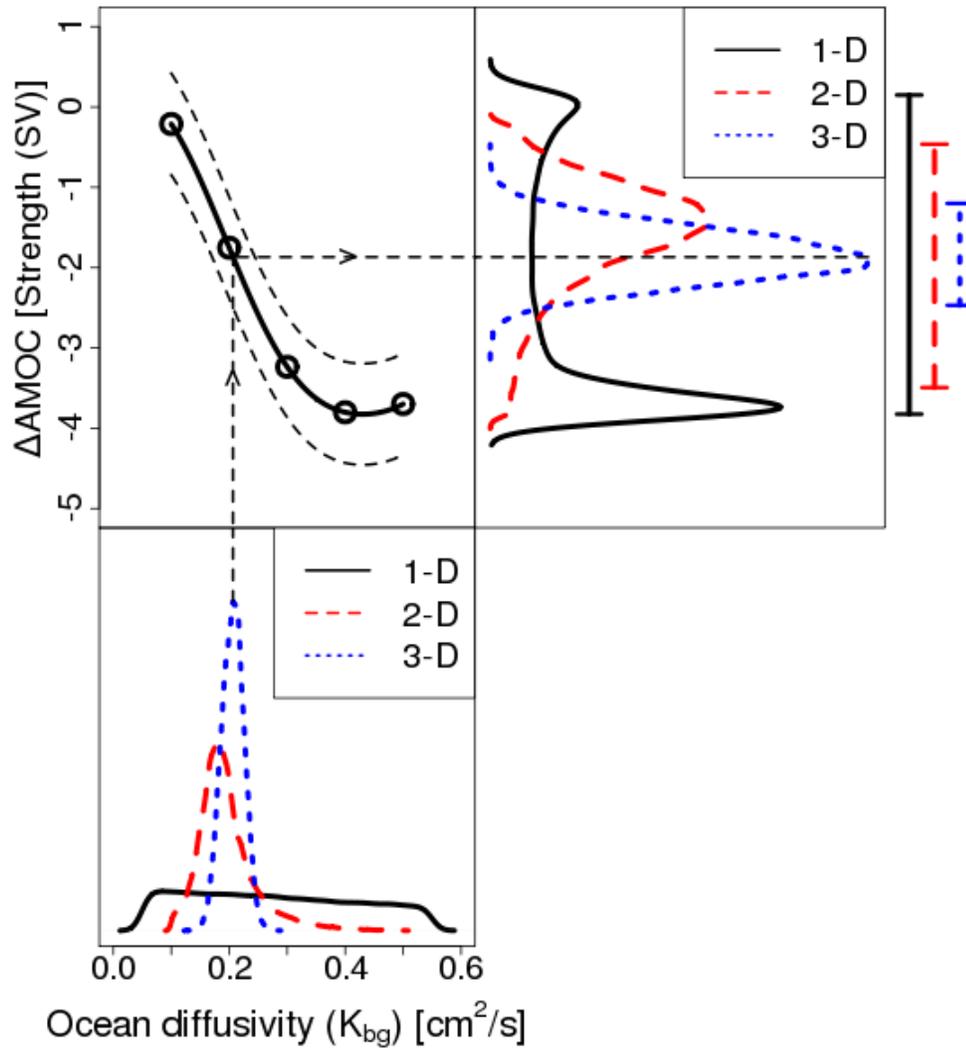


Figure 4.5: Combined posterior densities of K_{bg} from different prior specifications (lower left), the relationship between K_{bg} and projected AMOC change of the 2070 - 2099 mean from the 1970 - 1999 mean (upper left), and the resulting AMOC change projections (upper right) using 1-D (solid black line), 2-D (dashed red line) and 3-D (dotted blue line) data with their 95% credible intervals (bars at the right).

Comparison for $K_{bg}=0.2$, $A_{sc}=1.5$, $C_s=3.11$

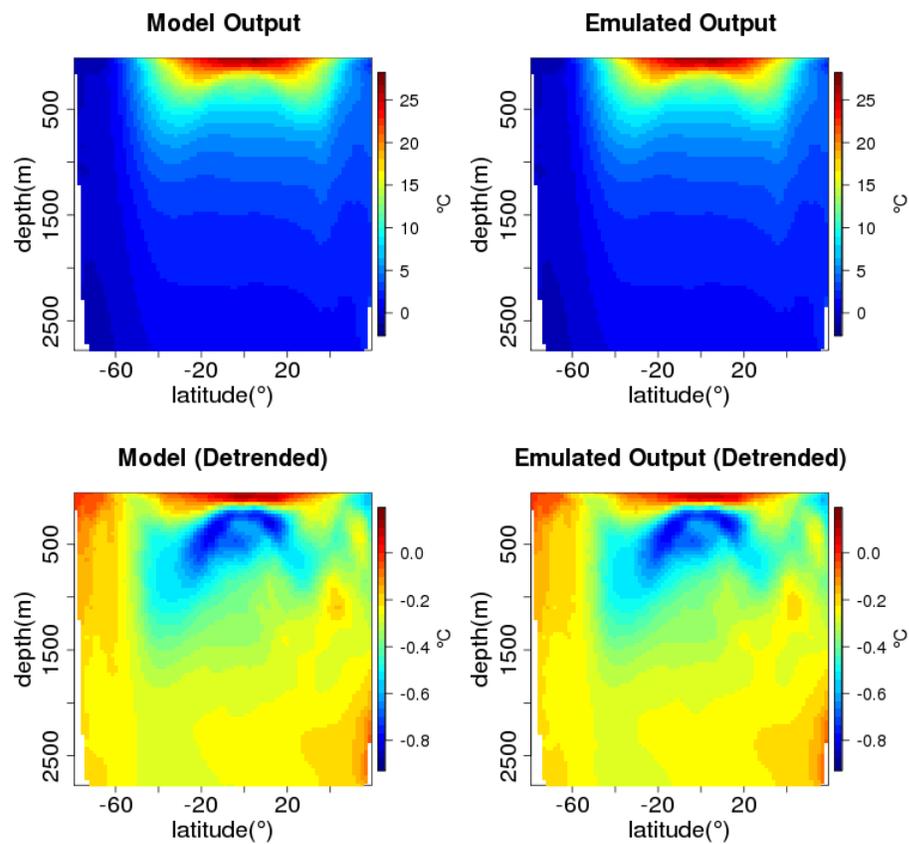


Figure 4.6: An example of our leave-10-percent out cross validation experiment for 2-D case (latitude-depth patterns). The first row shows the comparison between the raw original output and the raw emulated output. The second row shows the same comparison using detrended outputs, which are computed by subtracting the mean across the parameter settings at each location.

Computer Model Calibration Using Block Composite Likelihood*

In the previous chapter we have introduced a computationally efficient approach using dimension reduction based on PCA and kernel convolution. Here we introduce a composite likelihood-based approach to perform computer model calibration with high-dimensional spatial data. While composite likelihood has been studied extensively in the context of spatial statistics, computer model calibration using composite likelihood poses several new challenges. We propose a computationally efficient approach for Bayesian computer model calibration using composite likelihood. We also develop a methodology based on asymptotic theory for adjusting the composite likelihood posterior distribution so that it accurately represents posterior uncertainties. We study the application of our approach in the context of calibration for a climate model.

5.1 Introduction

Complex computer models are often used to approximate real-world processes. These models enable us to conduct virtual experiments that are useful for studying and understanding complex physical phenomena. A central source of uncertainty regarding computer models, and hence the behavior of the process they are ap-

*This chapter is based on a manuscript accepted for publication in *Statistica Sinica*, authored by Won chang, Murali Haran, Roman Olson, and Klaus Keller.

proximating, stems from uncertainty about the value of model input parameters. It is, however, often possible to learn about model parameter values from observations of the system being modeled. Computer model calibration, the methods used to learn about these parameters, involves finding model parameter settings that produce computer model outputs that are most compatible with the observed realization of the process. Statistical computer model calibration is a formal approach to parameter inference based on observations and on computer model output at various parameter settings. A sound approach to computer model calibration accounts for various sources of uncertainties such as measurement error and model structural errors, and results in a probability distribution that summarizes our knowledge about the parameters. Quantifying uncertainties about the parameters carefully is important as this allows for a rigorous quantification of uncertainties about projections based on the model. Here we consider computer model calibration for problems where the observations and the model output are in the form of spatial data.

Computer model calibration can pose nontrivial inferential challenges. In many applications computer model runs are computationally expensive. In this case, model runs are often available at only a limited number of parameter settings. A popular method to overcome this hurdle is the Gaussian process approach (cf. Sacks et al., 1989; Kennedy and O’Hagan, 2001). This method enables calibration with a limited number of model runs using probabilistic interpolation between the model runs. However, this approach faces computational challenges when applied to computer model output that are in the form of high-dimensional spatial data, which are increasingly common in modern science and engineering applications (see, e.g. Higdon et al., 2009; Bhat et al., 2010, 2012; Chang et al., 2014).

Some approaches have been developed recently to resolve these computational issues (e.g. Bayarri et al., 2007; Higdon et al., 2008; Bhat et al., 2012; Chang et al., 2014). Here we propose a new Bayesian approach for calibration with high-dimensional spatial data using composite likelihood methods. To our knowledge this is the first time composite likelihood methodology has been used in this context. This approach provides a way to mitigate the computational challenges for a wide range of computer model calibration problems.

The basic idea of composite likelihood (Besag, 1975, 1977; Lindsay, 1988) is

to approximate the original likelihood as a product of computationally cheaper likelihoods. This approach can be easily adapted for spatial modeling in various ways such as conditional likelihood (Vecchia, 1988; Stein et al., 2004), pairwise likelihood (Heagerty and Lele, 1998; Curriero and Lele, 1999; Ribatet et al., 2012), and block likelihood (Caragea and Smith, 2006; Eidsvik et al., 2013). Here we construct a calibration method based on block composite likelihood. In particular, we adopt the idea of hybrid composite likelihood proposed by Caragea and Smith (2006) that relies on two components: dependence between block means, and dependence within each block conditioning on its block mean. This composite likelihood approach allows for a substantial reduction in the computational burden for maximum likelihood inference with high-dimensional spatial data. Also, this opens up possibilities for flexible spatial covariance structure that vary depending on each block. Moreover, since the composite likelihood from the block composite likelihood framework is a valid probability model, no further justification is necessary for its use in Bayesian inference.

The remainder of this chapter is organized as follows. In Section 5.2 we outline the basic model calibration framework using Gaussian random fields. In Section 5.3 we formulate the Bayesian calibration model using block composite likelihood, discuss relevant asymptotic theory, and explain how Godambe information may be used to adjust posterior uncertainty when using composite likelihood. In Section 5.4 we describe an application of our method to a climate model calibration problem using 2-dimensional spatial patterns of ocean temperature change and a relevant simulated example. Finally, in Section 5.5, we conclude with a discussion and future directions for research.

5.2 Calibration Using Gaussian Processes

Here we introduce our computer model calibration framework which consists of two stages: model emulation and parameter calibration (Bayarri et al., 2007; Bhat et al., 2012; Chang et al., 2014). We first construct an ‘emulator’, which is a statistical model interpolating the computer model outputs as well as providing interpolation uncertainties (Sacks et al., 1989). Using the emulator, we find the posterior density of computer model parameters while taking into account impor-

tant sources of uncertainty including interpolation uncertainty, model-observation discrepancy, and observational error (Kennedy and O’Hagan, 2001).

We use the following notation henceforth. $Y(\mathbf{s}, \boldsymbol{\theta})$ is the computer model output at the spatial location $\mathbf{s} \in \mathcal{S}$ and the parameter setting $\boldsymbol{\theta} \in \Theta$. \mathcal{S} is the spatial field that we are interested in, usually a subset of \mathbb{R}^2 or \mathbb{R}^3 . $\Theta \subset \mathbb{R}^q$ is the open set of all possible computer model parameter settings with an integer $q \geq 1$. Let $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p\} \subset \Theta$ be a collection of p design points in the parameter space and $\{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subset \mathcal{S}$ be the set of n model grid locations. $\mathbf{Y}_i = (Y(\mathbf{s}_1, \boldsymbol{\theta}_i), \dots, Y(\mathbf{s}_n, \boldsymbol{\theta}_i))^T$ is computer model output at the model grid locations at the parameter setting $\boldsymbol{\theta}_i$. The concatenated $np \times 1$ vector of all computer model outputs is $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_p^T)^T$. Typically $p \ll n$ since computer model runs with high-resolution are computationally expensive. Finally, we let $Z(\mathbf{s})$ be an observation at spatial location \mathbf{s} and $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$ be the observational data, a spatial process observed at n locations. The notation introduced in this section is summarized in Table 5.1.

Model Emulation Using Gaussian Processes. Following Bhat et al. (2012)

Table 5.1: Summary of notation used in Section 5.2

Symbol	Definition
p	number of design points in parameter space
n	number of spatial locations for computer model grid
Θ	open set of all possible computer model parameter settings
$\boldsymbol{\theta}_i$	parameter setting for i th model run
\mathcal{S}	spatial field of interest
\mathbf{s}_j	j th location on computer model grid
$Y(\mathbf{s}_j, \boldsymbol{\theta}_i)$	model output at location \mathbf{s}_j for parameter setting $\boldsymbol{\theta}_i$
\mathbf{Y}_i	model output for i th parameter setting, $(Y(\mathbf{s}_1, \boldsymbol{\theta}_i), \dots, Y(\mathbf{s}_n, \boldsymbol{\theta}_i))^T$
$Z(\mathbf{s}_j)$	observation at spatial location \mathbf{s}_j
\mathbf{Z}	observational data, $(Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$
\mathbf{X}	covariate matrix
$\boldsymbol{\beta}$	vector of regression coefficients
$\boldsymbol{\xi}_y$	vector of covariance parameters for emulator
$\eta(\mathbf{s}, \boldsymbol{\theta})$	emulator process at location \mathbf{s} and $\boldsymbol{\theta}$
$\boldsymbol{\eta}(\boldsymbol{\theta})$	emulator output at parameter setting $\boldsymbol{\theta}$, $(\eta(\mathbf{s}_1, \boldsymbol{\theta}), \dots, \eta(\mathbf{s}_n, \boldsymbol{\theta}))^T$
$\boldsymbol{\theta}^*$	true or fitted value of computer model parameter for observational data
$\delta(\mathbf{s})$	model-observation discrepancy at location \mathbf{s}
$\boldsymbol{\delta}$	discrepancy process observed at grid locations, $(\delta(\mathbf{s}_1), \dots, \delta(\mathbf{s}_n))^T$

and Chang et al. (2014), we construct a Gaussian process that interpolates computer model outputs as

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

where \mathbf{X} is an $np \times b$ covariate matrix containing all the spatial locations and climate parameters used to define the covariance matrix $\Sigma(\boldsymbol{\xi}_y)$. $\boldsymbol{\beta}$ and $\boldsymbol{\xi}_y$ are the vectors of regression coefficients and covariance parameters respectively. We construct an interpolation process by finding the maximum likelihood estimate (MLE) of these parameters. This interpolation model provides the predictive distribution of a computer model run at any given location $\mathbf{s} \in \mathcal{S}$ and $\boldsymbol{\theta} \in \Theta$ (Sacks et al., 1989). We call this predictive process an emulator and denote it by $\eta(\mathbf{s}, \boldsymbol{\theta})$. Throughout this chapter, $\boldsymbol{\beta}$ is set to $\mathbf{0}$ since the Gaussian process provides enough flexibility in modeling the output process.

Model Calibration Using Gaussian Random Processes. We model the observational data \mathbf{Z} as

$$\mathbf{Z} = \boldsymbol{\eta}(\boldsymbol{\theta}^*) + \boldsymbol{\delta}, \tag{5.1}$$

where $\boldsymbol{\theta}^*$ is the true or fitted value of computer model parameter for the observational data (Bayarri et al., 2007), $\boldsymbol{\eta}(\boldsymbol{\theta}^*) = (\eta(\mathbf{s}_1, \boldsymbol{\theta}^*), \dots, \eta(\mathbf{s}_n, \boldsymbol{\theta}^*))^T$ is the emulator output at $\boldsymbol{\theta}^*$ on the model grid, and $\boldsymbol{\delta} = (\delta(\mathbf{s}_1), \dots, \delta(\mathbf{s}_n))^T$ is a term that includes both data-model discrepancy as well as observational error. The discrepancy process $\delta(\mathbf{s})$ is also modeled as a Gaussian process with spatial covariance between the locations $\mathbf{s}_1, \dots, \mathbf{s}_n$. Model calibration with high-dimensional spatial data leads to computational challenges as described in the following section.

5.3 Calibration with High-Dimensional Spatial Data

In this section we briefly examine the challenges in model calibration using high-dimensional spatial data and the existing approaches to the problem. We then proceed to the formulation of our composite likelihood approach. We summarize the notation used in this section in Tables 5.2 and 5.3.

Table 5.2: Summary of notation used for emulation model in Section 5.3

Symbol	Definition
M	number of blocks
n_i	number of spatial locations in i th block
\mathbf{s}_{ij}	j th spatial location in i th block
$\mathbf{Y}_{(i)}$	model output for i th block, $(Y(\mathbf{s}_{i1}, \cdot)^T, Y(\mathbf{s}_{i2}, \cdot)^T, \dots, Y(\mathbf{s}_{in_i-1}, \cdot)^T)^T$
$Y(\mathbf{s}_{ij}, \cdot)$	model output at location \mathbf{s}_{ij} , $(Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_1), \dots, Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_p))^T$
$\bar{\mathbf{Y}}_{(i)}$	mean vector for i th block, $\frac{1}{n_i} \sum_{j=1}^{n_i} (Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_1), \dots, Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_p))^T$
$\bar{\mathbf{Y}}$	vector of all block means, $(\bar{\mathbf{Y}}_{(1)}^T, \dots, \bar{\mathbf{Y}}_{(M)}^T)$
$\boldsymbol{\xi}_s$	parameters for spatial covariance function K_s
$K_s(\mathbf{s}, \mathbf{s}'; \boldsymbol{\xi}_s)$	spatial covariance function between \mathbf{s} and \mathbf{s}' for emulator
$\boldsymbol{\xi}_\theta$	parameters for covariance function K_θ
$K_\theta(\boldsymbol{\theta}, \boldsymbol{\theta}'; \boldsymbol{\xi}_\theta)$	covariance function between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$
$\Sigma^{\bar{\mathbf{Y}}}$	covariance between the block means, $\mathbf{H} \otimes \Sigma_\theta$
\mathbf{H}	spatial covariance for block means, $\{\mathbf{H}\}_{ij} = \frac{1}{n_i n_j} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} K_s(\mathbf{s}_{ik}, \mathbf{s}_{jl}; \boldsymbol{\xi}_s)$
Σ_θ	covariance between p parameter settings, $\{\Sigma_\theta\}_{ij} = K_\theta(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j; \boldsymbol{\xi}_\theta)$
$\mu_i^{\mathbf{Y} \bar{\mathbf{Y}}}$	conditional mean of $\mathbf{Y}_{(i)}$ given $\bar{\mathbf{Y}}_{(i)}$, $(\gamma^{(i)} / \{\mathbf{H}\}_{ii} \otimes \mathbf{I}_p) \bar{\mathbf{Y}}_{(i)}$
$\Sigma_i^{\mathbf{Y} \bar{\mathbf{Y}}}$	conditional variance of $\mathbf{Y}_{(i)}$ given $\bar{\mathbf{Y}}_{(i)}$, $(\Gamma_i - \gamma^{(i)} (\gamma^{(i)})^T / \{\mathbf{H}\}_{ii}) \otimes \Sigma_\theta$
$\gamma^{(i)}$	spatial covariance between i th block mean and block locations, $\{\gamma^{(i)}\}_j = \frac{1}{n_i} \sum_{k=1}^{n_i} K_s(\mathbf{s}_{ij}, \mathbf{s}_{ik}; \boldsymbol{\xi}_s)$
Γ_i	spatial covariance matrix for i th block, $\{\Gamma_i\}_{jk} = K_s(\mathbf{s}_{ij}, \mathbf{s}_{ik}; \boldsymbol{\xi}_s)$

5.3.1 Challenges with High-Dimensional Spatial Data

The basic challenge with the approach in Section 5.2 stems from the fact that the computational cost for a single likelihood evaluation is $\mathcal{O}(n^3 p^3)$. For large n , evaluating the likelihood function repeatedly when using algorithms like Markov chain Monte Carlo (MCMC) can become computationally prohibitive. One can reduce the computational cost by assuming a separable covariance structure between the spatial dependence and the dependence due to computer model parameters, but the computational cost is still $\mathcal{O}(n^3)$, and hence does not scale well with n . The current approaches to overcome such limitation for high-dimensional data rely on dimension reduction or basis expansion. The dimension reduction approaches (Bayarri et al., 2007; Chang et al., 2014) map the original output into a lower dimension and exploit the uncorrelated nature of the low-dimensional processes to speed up the computation. The basis expansion approaches (Bhat et al., 2012; Higdon et al., 2008) use a basis representation of model output that results in a reformulated likelihood with a lower computational cost. Here we introduce

Table 5.3: Summary of notation used for calibration model in Section 5.3

Symbol	Definition
$\mathbf{Z}_{(i)}$	observations for i th block, $(Z(\mathbf{s}_{i1}), \dots, Z(\mathbf{s}_{in_i}))^T$
$\bar{\mathbf{Z}}_{(i)}$	i th block mean for observations
$\bar{\mathbf{Z}}$	collection of block means for observations, $(\bar{\mathbf{Z}}_{(1)}, \dots, \bar{\mathbf{Z}}_{(M)})^T$
ξ_d	parameters for spatial covariance function K_d
$K_d(\mathbf{s}, \mathbf{s}'; \xi_d)$	spatial covariance function between \mathbf{s} and \mathbf{s}' for discrepancy
Ω	spatial covariance matrix for block means of discrepancy, $\{\Omega\}_{ij} = \frac{1}{n_i n_j} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} K_d(\mathbf{s}_{ik}, \mathbf{s}_{jl}; \xi_d)$
$\Sigma_{\theta^* \theta}$	covariance between θ^* and θ
$\mu^{\bar{\mathbf{Z}}}$	mean vector for the block means $\bar{\mathbf{Z}}$, $(\mathbf{I}_M \otimes \Sigma_{\theta^* \theta} \Sigma_{\theta}^{-1}) \bar{\mathbf{Y}}$
$\Sigma^{\bar{\mathbf{Z}}}$	covariance matrix for $\bar{\mathbf{Z}}$, $\mathbf{H} \otimes (\Sigma_{\theta^*} - \Sigma_{\theta^* \theta} \Sigma_{\theta}^{-1} \Sigma_{\theta^* \theta}^T) + \Omega$
$\lambda^{(i)}$	spatial covariance between i th block mean and block locations for discrepancy, $\{\lambda^{(i)}\}_j = \frac{1}{n_i} \sum_{k=1}^{n_i} K_d(\mathbf{s}_{ij}, \mathbf{s}_{ik}; \xi_d)$
Λ_i	spatial covariance for i th block locations for discrepancy, $\{\Lambda_i\}_{jk} = K_d(\mathbf{s}_{ij}, \mathbf{s}_{ik}; \xi_d)$
$\mu_i^{\mathbf{Z} \bar{\mathbf{Z}}}$	mean of $\mathbf{Z}_{(i)}$ given $\bar{\mathbf{Z}}_{(i)}$, $(\mathbf{I}_{n_i-1} \otimes \Sigma_{\theta^* \theta} \Sigma_{\theta}^{-1}) \mathbf{Y}_{(i)} + (\tau^{(i)} + \lambda^{(i)}) \{\Sigma^{\bar{\mathbf{Z}}}\}_{ii}^{-1} (\bar{\mathbf{Z}}_{(i)} - \{\mu^{\bar{\mathbf{Z}}}\}_i)$
$\Sigma_i^{\mathbf{Z} \bar{\mathbf{Z}}}$	covariance of $\mathbf{Z}_{(i)}$ given $\bar{\mathbf{Z}}_{(i)}$, $(\Gamma_i \otimes (\Sigma_{\theta^*} - \Sigma_{\theta^* \theta} \Sigma_{\theta}^{-1} \Sigma_{\theta^* \theta}^T) + \Lambda_i) - (\tau^{(i)} + \lambda^{(i)})(\tau^{(i)} + \lambda^{(i)})^T / \{\Sigma^{\bar{\mathbf{Z}}}\}_{ii}$
$\tau^{(i)}$	$\gamma^{(i)} \otimes (\Sigma_{\theta^*} - \Sigma_{\theta^* \theta} \Sigma_{\theta}^{-1} \Sigma_{\theta^* \theta}^T)$

a somewhat different approach that relies on the block composite likelihood for spatial data (Caragea and Smith, 2006; Eidsvik et al., 2013).

5.3.2 Composite Likelihood for Model Calibration

In this framework, we partition the spatial field \mathcal{S} into small blocks to avoid the computational issues related to high-dimensional data. In Section 5.4 we describe an example of how such a partition can be constructed in practice. The block composite likelihood method substitutes the original likelihood by a composite likelihood that utilizes the spatial blocks, thereby resulting in a likelihood function that requires much less computational effort. In particular, we adopt the block composite likelihood formulation of Caragea and Smith (2006). This framework assumes conditional independence between outcomes in different blocks given the block means, and the dependence between blocks is modeled through the covariance between block means. This framework gives a valid probability model, and therefore the posterior distribution defined using the composite likelihood function based on this approach is also a valid probability model. Obtaining a valid prob-

ability model is important because we are embedding the likelihood in a Bayesian approach; having a valid probability model automatically assures us that the resulting posterior distribution is proper when all the prior distributions used are proper.

We do not use spatial blocking to fully specify the probability model here. Instead, the blocking provides a way to find a computationally feasible pseudo-likelihood estimator that has good asymptotic properties under the original probability model described in Section 5.2. Since the composite likelihood estimator based on blocking is not the correct likelihood estimator under the original probability model, an additional adjustment step for its asymptotic variance is required and this is discussed below.

We divide the spatial area for the computer model output into M different blocks and denote the output for each block by $\mathbf{Y}_{(1)}, \dots, \mathbf{Y}_{(M)}$. The blocks are made according to the spatial field, not the parameter space, because the number of computer model runs is usually quite limited due to the high computational costs of running the model. However, in principle our approach can be extended to blocking in parameter space as well if the number of model runs is also large. Let n_i denote the number of computer model outcomes in the i th block. We denote the spatial locations in the i th block by $\mathbf{s}_{i1}, \dots, \mathbf{s}_{in_i}$. Each $\mathbf{Y}_{(i)}$ is a stack of $(n_i - 1)$ -dimensional spatial output for p different parameter settings,

$$\mathbf{Y}_{(i)} = \left(Y(\mathbf{s}_{i1}, \cdot)^T, Y(\mathbf{s}_{i2}, \cdot)^T, \dots, Y(\mathbf{s}_{in_i-1}, \cdot)^T \right)^T,$$

where $Y(\mathbf{s}_{ij}, \cdot) = (Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_1), \dots, Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_p))^T$ is the $p \times 1$ vector of computer model outcomes for all the parameter settings $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p$. We omit one spatial location for each block in defining the output vectors to avoid degeneracy. The composite likelihood function changes slightly depending on which spatial location in each block is omitted. However, this difference is negligible unless the block sizes n_1, \dots, n_M are very small. We let $\bar{\mathbf{Y}}_{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_1), \dots, Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_p))^T$ be the p -dimensional mean vector of model outcomes for the i th block, the means for the spatial block consisting of same set of locations across all model parameter settings. We define the vector of all block means by $\bar{\mathbf{Y}} = \left(\bar{\mathbf{Y}}_{(1)}^T, \dots, \bar{\mathbf{Y}}_{(M)}^T \right)^T$. Similarly, we divide the observational data into M blocks in the same way and omit one observation

for each block to have $\mathbf{Z}_{(1)}, \dots, \mathbf{Z}_{(M)}$, the vectors of observational data in different blocks. We let $\bar{\mathbf{Z}}_{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} Z(\mathbf{s}_{ij})$ be the i th block mean of observational data and $\bar{\mathbf{Z}} = (\bar{\mathbf{Z}}_{(1)}, \dots, \bar{\mathbf{Z}}_{(M)})^T$ be the collection of them.

Assuming separability, we model the covariance between the process at two different spatial locations and parameter settings $Y(\mathbf{s}, \boldsymbol{\theta})$ and $Y(\mathbf{s}', \boldsymbol{\theta}')$ by

$$\text{Cov}(Y(\mathbf{s}, \boldsymbol{\theta}), Y(\mathbf{s}', \boldsymbol{\theta}')) = K_s(\mathbf{s}, \mathbf{s}'; \boldsymbol{\xi}_s) K_\theta(\boldsymbol{\theta}, \boldsymbol{\theta}'; \boldsymbol{\xi}_\theta),$$

where K_s and K_θ are valid covariance functions respectively in \mathcal{S} and Θ with parameters $\boldsymbol{\xi}_s$ and $\boldsymbol{\xi}_\theta$. The covariance between discrepancy process at \mathbf{s} and \mathbf{s}' is given by

$$\text{Cov}(\delta(\mathbf{s}), \delta(\mathbf{s}')) = K_d(\mathbf{s}, \mathbf{s}'; \boldsymbol{\xi}_d),$$

with a valid covariance function K_d in \mathcal{S} and a vector of parameters $\boldsymbol{\xi}_d$. More specific definition of the covariance functions is discussed below.

Computer Model Emulation. The first component of our composite likelihood is the model for block means that captures the large scale trend. The covariance between the block means is $\Sigma^{\bar{\mathbf{Y}}} = \mathbf{H} \otimes \Sigma_\theta$ where Σ_θ is the covariance matrix for the random variable across p parameter settings and \mathbf{H} is the $M \times M$ covariance matrix between the blocks. It is straightforward to see that the block covariance is

$$\{\mathbf{H}\}_{ij} = \frac{1}{n_i n_j} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} K_s(\mathbf{s}_{ik}, \mathbf{s}_{jl}; \boldsymbol{\xi}_s), \quad (5.2)$$

the mean of all possible cross covariances between two blocks.

The second component is the sum of the conditional likelihoods for each block, which models the small scale dependence and variation. For the i th block, the conditional distribution of output $\mathbf{Y}_{(i)}$ given the block mean $\bar{\mathbf{Y}}_{(i)}$ is a normal distribution with the mean and covariance given by

$$\begin{aligned} \mu_i^{\mathbf{Y}|\bar{\mathbf{Y}}} &= E(\mathbf{Y}_{(i)}|\bar{\mathbf{Y}}_{(i)}) = (\gamma^{(i)} / \{\mathbf{H}\}_{ii} \otimes \mathbf{I}_p) \bar{\mathbf{Y}}_{(i)} \\ \Sigma_i^{\mathbf{Y}|\bar{\mathbf{Y}}} &= \text{Var}(\mathbf{Y}_{(i)}|\bar{\mathbf{Y}}_{(i)}) = (\Gamma_i - \gamma^{(i)} (\gamma^{(i)})^T / \{\mathbf{H}\}_{ii}) \otimes \Sigma_\theta, \end{aligned}$$

where Γ_i is the $(n_i - 1) \times (n_i - 1)$ spatial covariance matrix for the i th block and $\gamma^{(i)}$ is the $(n_i - 1) \times 1$ covariance vector between the i th block mean and the i th

block locations (see Table 5.2 for their definitions). The log composite likelihood function for the model output is then

$$\begin{aligned} \text{cl}(\boldsymbol{\xi}_s, \boldsymbol{\xi}_\theta) \propto & -\frac{1}{2} \left(\log |\Sigma^{\bar{\mathbf{Y}}}| + \bar{\mathbf{Y}}^T (\Sigma^{\bar{\mathbf{Y}}})^{-1} \bar{\mathbf{Y}} \right) \\ & - \frac{1}{2} \sum_{i=1}^M \left(\log |\Sigma_i^{\mathbf{Y}|\bar{\mathbf{Y}}}| + (\mathbf{Y}_{(i)} - \mu_i^{\mathbf{Y}|\bar{\mathbf{Y}}})^T (\Sigma_i^{\mathbf{Y}|\bar{\mathbf{Y}}})^{-1} (\mathbf{Y}_{(i)} - \mu_i^{\mathbf{Y}|\bar{\mathbf{Y}}}) \right). \end{aligned}$$

We construct the emulator by finding the MLE of $\boldsymbol{\xi}_\theta$ and $\boldsymbol{\xi}_s$, denoted by $\hat{\boldsymbol{\xi}}_\theta$ and $\hat{\boldsymbol{\xi}}_s$. The computational cost for a single likelihood evaluation is reduced from $\frac{1}{3}n^3$ flops to $\sum_{i=1}^M \sum_{j=i}^M n_i n_j + \frac{1}{3}(M^3) + \frac{1}{3} \left(\sum_{i=1}^M (n_i - 1)^3 \right)$ flops, where the first term is the computational cost for finding H. This is a reduction from 6.86×10^{10} flops to 5.92×10^7 flops in the climate model calibration example in Section 5.4.

Computer Model Calibration. We formulate the composite likelihood for observational data in the same manner as above. Let $\boldsymbol{\Omega}$ be the $M \times M$ covariance between the M block means of the discrepancy $\boldsymbol{\delta}$, defined in the same way as H, but with a different set of parameters $\boldsymbol{\xi}_d$ (see Table 5.3 for its definition). The conditional mean and covariance for the block means of observational data $\bar{\mathbf{Z}}$ are

$$\begin{aligned} \mu^{\bar{\mathbf{Z}}} &= (\mathbf{I}_M \otimes \Sigma_{\theta^*} \Sigma_\theta^{-1}) \bar{\mathbf{Y}}, \text{ an } M \times 1 \text{ vector,} \\ \Sigma^{\bar{\mathbf{Z}}} &= \text{H} \otimes (\Sigma_{\theta^*} - \Sigma_{\theta^*} \Sigma_\theta^{-1} \Sigma_{\theta^*}^T) + \boldsymbol{\Omega}, \text{ an } M \times M \text{ matrix.} \end{aligned}$$

where Σ_{θ^*} is the $1 \times p$ covariance vector between the fitted computer model parameter value $\boldsymbol{\theta}^*$ and the design points $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p$. Likewise, we define Λ_i and $\lambda^{(i)}$ as the discrepancy counterparts of Γ_i and $\gamma^{(i)}$ with the covariance parameter $\boldsymbol{\xi}_d$. Hence, Λ_i and $\lambda^{(i)}$ are the $(n_i - 1) \times (n_i - 1)$ i th block discrepancy covariance matrix and the $(n_i - 1) \times 1$ covariance vector between the block outputs and the block mean respectively (see Table 5.3 for their definitions). The conditional mean and covariance for observational data in the i th block are therefore

$$\begin{aligned} \mu_i^{\mathbf{Z}|\bar{\mathbf{Z}}} &= (\mathbf{I}_{n_i-1} \otimes \Sigma_{\theta^*} \Sigma_\theta^{-1}) \mathbf{Y}_{(i)} + (\tau^{(i)} + \lambda^{(i)}) \left\{ \Sigma^{\bar{\mathbf{Z}}} \right\}_{ii}^{-1} (\bar{\mathbf{Z}}_i - \left\{ \mu^{\bar{\mathbf{Z}}} \right\}_i), \\ \Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}} &= (\Gamma_i \otimes (\Sigma_{\theta^*} - \Sigma_{\theta^*} \Sigma_\theta^{-1} \Sigma_{\theta^*}^T) + \Lambda_i) - (\tau^{(i)} + \lambda^{(i)}) (\tau^{(i)} + \lambda^{(i)})^T / \left\{ \Sigma^{\bar{\mathbf{Z}}} \right\}_{ii}, \end{aligned}$$

where $\tau^{(i)} = \gamma^{(i)} \otimes (\Sigma_{\theta^*} - \Sigma_{\theta^*} \Sigma_\theta^{-1} \Sigma_{\theta^*}^T)$. The log composite likelihood for the

observational data is then

$$\begin{aligned} c\ell_n(\boldsymbol{\psi}) \propto & -\frac{1}{2} \left(\log |\Sigma^{\bar{\mathbf{z}}}| + (\bar{\mathbf{z}} - \mu^{\bar{\mathbf{z}}})^T (\Sigma^{\bar{\mathbf{z}}})^{-1} (\bar{\mathbf{z}} - \mu^{\bar{\mathbf{z}}}) \right) \\ & - \frac{1}{2} \sum_{i=1}^M \left(\log |\Sigma_i^{\mathbf{z}|\bar{\mathbf{z}}}| + (\mathbf{z}_{(i)} - \mu_i^{\mathbf{z}|\bar{\mathbf{z}}})^T (\Sigma_i^{\mathbf{z}|\bar{\mathbf{z}}})^{-1} (\mathbf{z}_{(i)} - \mu_i^{\mathbf{z}|\bar{\mathbf{z}}}) \right), \end{aligned} \quad (5.3)$$

where the first line in (5.3) is the log likelihood corresponding to the block means and the second line corresponding to the observations within each block. $\boldsymbol{\psi}$ denotes all the parameters being estimated in the calibration stage including $\boldsymbol{\theta}^*$ and $\boldsymbol{\xi}_d$.

By choosing a proper prior for $\boldsymbol{\psi}$, $f(\boldsymbol{\psi})$, we define the approximate log posterior density, $\log(\pi_n(\boldsymbol{\psi})) \propto \log f(\boldsymbol{\psi}) + c\ell_n(\boldsymbol{\psi})$ and infer $\boldsymbol{\psi}$ using the standard Metropolis-Hastings algorithm. We allow the scale parameters for the emulator to be re-estimated along with the other parameters but fix the other emulator parameters in $\boldsymbol{\xi}_s$ and $\boldsymbol{\xi}_\theta$ at their estimated values from the emulation stage (Bayarri et al., 2007; Bhat et al., 2012; Chang et al., 2014). The formulation results in the same computational gain as in the emulation stage.

In both the emulation and calibration stages, calculation of the covariance matrix for the block means is a computational bottleneck, requiring $\sum_{i=1}^M \sum_{j=i}^M n_i n_j$ flops of computation. While computationally very demanding, its contribution to the likelihood function is usually not significant (Caragea and Smith, 2006). Therefore, instead of using all cross covariances between spatial locations, we randomly sample a subset of cross covariances to approximate the covariance between block means \mathbf{H} . The computation of \mathbf{H} in (5.2) is substituted by

$$\{\mathbf{H}\}_{ij} = \frac{1}{m_i m_j} \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} K_s(\mathbf{u}_{ik}, \mathbf{u}_{jl}; \boldsymbol{\xi}_s), \quad (5.4)$$

with $m_i \leq n_i$ and $m_j \leq n_j$, where $\mathbf{u}_{i1}, \dots, \mathbf{u}_{im_i}$ and $\mathbf{u}_{j1}, \dots, \mathbf{u}_{jm_j}$ are randomly chosen respectively from $\mathbf{s}_{i1}, \dots, \mathbf{s}_{in_i}$ and $\mathbf{s}_{j1}, \dots, \mathbf{s}_{jn_j}$. This reduces the computational cost from $\sum_{i=1}^M \sum_{j=i}^M n_i n_j$ to $\sum_{i=1}^M \sum_{j=i}^M m_i m_j$, that is, 1.32×10^7 flops to 2.86×10^5 flops for the calibration problem in Section 5.4.2. The same approximation can be applied to Ω with $\boldsymbol{\xi}_d$.

Covariance Function and Prior Specification. We use the exponential covariance function with a nugget term to define the covariance between parameter

settings (K_θ), spatial covariance for the emulator (K_s), and the spatial covariance for the discrepancy (K_d). Thus the covariance between the process at two parameter settings $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^T$ and $\boldsymbol{\theta}' = (\theta'_1, \dots, \theta'_q)^T$ is

$$K_\theta(\boldsymbol{\theta}, \boldsymbol{\theta}'; \boldsymbol{\xi}_\theta) = \zeta_\theta 1(\boldsymbol{\theta} = \boldsymbol{\theta}') + \kappa_\theta \exp\left(-\sum_{i=1}^q \phi_{\theta,i} |\theta_i - \theta'_i|\right),$$

where $\boldsymbol{\xi}_\theta = (\zeta_\theta, \kappa_\theta, \phi_{\theta,1}, \dots, \phi_{\theta,q})$, and $\zeta_\theta, \kappa_\theta, \phi_{\theta,1}, \dots, \phi_{\theta,q} > 0$. Likewise, the covariance between the process at two spatial locations \mathbf{s} and \mathbf{s}' for the emulator and the discrepancy term are

$$\begin{aligned} K_s(\mathbf{s}, \mathbf{s}'; \boldsymbol{\xi}_s) &= \kappa_s (\zeta_s 1(\mathbf{s} = \mathbf{s}') + \exp(-\phi_s g(\mathbf{s}, \mathbf{s}'))), \\ K_d(\mathbf{s}, \mathbf{s}'; \boldsymbol{\xi}_d) &= \kappa_d (\zeta_d 1(\mathbf{s} = \mathbf{s}') + \exp(-\phi_d g(\mathbf{s}, \mathbf{s}'))), \end{aligned} \quad (5.5)$$

respectively, with $\boldsymbol{\xi}_s = (\zeta_s, \kappa_s, \phi_s)$, $\boldsymbol{\xi}_d = (\zeta_d, \kappa_d, \phi_d)$, and $\zeta_s, \kappa_s, \phi_s, \zeta_d, \kappa_d, \phi_d > 0$. $g(\mathbf{s}, \mathbf{s}')$ denotes the distance between two points. In the climate model calibration problem in Section 5.4, for example, g is the geodesic distance between two points on the Earth's surface.

The parameters inferred by the Bayesian approach in the calibration stage are κ_s , ζ_d , κ_d , ϕ_d , and $\boldsymbol{\theta}^*$. Following Bayarri et al. (2007), the sill parameter for the emulator κ_s is initially inferred via maximum likelihood estimate in the emulation stage and re-estimated by Bayesian inference in the calibration stage. We impose informative priors on the above parameters to avoid potentially obtaining improper posterior distributions (cf. Berger et al., 2001) and identifiability issues. The latter is explained further in Section 5.4. The sill parameters, κ_s and κ_d receive inverse-Gamma priors $IG(a_{\kappa_s}, b_{\kappa_s})$ and $IG(a_{\kappa_d}, b_{\kappa_d})$. We also impose an Inverse-Gamma prior $IG(a_{\zeta_d}, b_{\zeta_d})$ for the nugget parameter ζ_d . The prior density for the range parameter ϕ_d is assumed to be uniform with a wide support. The fitted computer model parameter $\boldsymbol{\theta}^*$ also receives a uniform prior over a wide range. Note that one can also assume a more informative prior for $\boldsymbol{\theta}^*$ such as a unimodal distribution based on some physical knowledge. However, in the calibration problem in Section 5.4 we do not impose such a prior for $\boldsymbol{\theta}^*$; this allows us to study the characteristics of the posterior density of $\boldsymbol{\theta}^*$ more transparently.

Asymptotics and Adjustment Using Godambe Information. The com-

posite likelihood in (5.3) is not based on the true probability model in (5.1), and therefore the ‘composite’ posterior density based on (5.3) is quite different from the true posterior based on (5.1). In this section, we discuss how the Godambe information matrix (Godambe, 1960) for estimating equations may be used to adjust for using the composite likelihood when making inferences.

We first provide the asymptotic justification for the adjustment using the Godambe information matrix. We show that, for large n and p , the mode of the approximate posterior $\hat{\boldsymbol{\psi}}_n^B = \arg \max_{\boldsymbol{\psi}} \pi_n(\boldsymbol{\psi})$ is consistent and asymptotically normally distributed, with a covariance matrix given by the inverse of the Godambe information matrix. If we let $p \rightarrow \infty$, then the emulator converges to the measurement-error model such that

$$\boldsymbol{\eta}(\boldsymbol{\theta}) \sim N(\mathbf{Y}(\boldsymbol{\theta}), \zeta_{\theta} \Sigma^s),$$

where $\mathbf{Y}(\boldsymbol{\theta})$ is the $n \times 1$ vector of model output at the parameter setting $\boldsymbol{\theta}$ and the spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_n$. This result holds as long as the computer model output varies reasonably smoothly in the parameter space (Yakowitz and Szidarovszky, 1985). The model for observational data is

$$\mathbf{Z} \sim N(\mathbf{Y}^*, \zeta_{\theta} \Sigma^s + \Sigma^d), \quad (5.6)$$

where $\mathbf{Y}^* = \mathbf{Y}(\boldsymbol{\theta}^*)$, $\{\Sigma^s\}_{ij} = K_s(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\xi}_s)$, and $\{\Sigma^d\}_{ij} = K_d(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\xi}_d)$. The composite likelihood in (5.3) then has means and covariances

$$\begin{aligned} \mu^{\bar{\mathbf{Z}}} &= \bar{\mathbf{Y}}^*, \text{ an } M \times 1 \text{ vector,} \\ \Sigma^{\bar{\mathbf{Z}}} &= \zeta_{\theta} \mathbf{H} + \Omega, \text{ an } M \times M \text{ matrix,} \\ \mu_i^{\mathbf{Z}|\bar{\mathbf{Z}}} &= \mathbf{Y}_{(i)}^* + (\zeta_{\theta} \gamma^{(i)} + \lambda^{(i)}) \left\{ \Sigma^{\bar{\mathbf{Z}}} \right\}_{ii}^{-1} (\bar{\mathbf{Z}}_i - \left\{ \mu^{\bar{\mathbf{Z}}} \right\}_i), \\ \Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}} &= (\zeta_{\theta} \Gamma_i + \Lambda_i) - (\zeta_{\theta} \gamma^{(i)} + \lambda^{(i)}) (\zeta_{\theta} \gamma^{(i)} + \lambda^{(i)})^T / \left\{ \Sigma^{\bar{\mathbf{Z}}} \right\}_{ii}, \end{aligned}$$

where $\bar{\mathbf{Y}}_{(i)}^* = \frac{1}{n_i} \sum_{j=1}^{n_i} Y(\mathbf{s}_{ij}, \boldsymbol{\theta}^*)$ is the i th block mean of the computer model output at $\boldsymbol{\theta}^*$ and $\bar{\mathbf{Y}}^* = \left(\bar{\mathbf{Y}}_{(1)}^*, \dots, \bar{\mathbf{Y}}_{(M)}^* \right)^T$ is the collection of all their block means.

We now show the consistency and the asymptotic normality of the posterior mode $\hat{\boldsymbol{\psi}}_n^B$ as $n \rightarrow \infty$. We utilize expanding domain asymptotic results (see, e.g.

Mardia and Marshall, 1984; Cressie, 1993; Cox and Reid, 2004; Zhang and Zimmerman, 2005; Varin, 2008). The first step is establishing consistency and asymptotic normality of the maximum composite likelihood estimator.

Proposition 5.3.1. *The following holds for the maximum composite likelihood estimator $\hat{\boldsymbol{\psi}}_n^{CL} = \arg \max_{\boldsymbol{\psi}} c\ell_n(\boldsymbol{\psi})$.*

(i) (Consistency) $\hat{\boldsymbol{\psi}}_n^{CL} \xrightarrow{\mathcal{P}} \boldsymbol{\psi}^0$ as $n \rightarrow \infty$, where $\boldsymbol{\psi}^0$ is the vector of true values of parameters in $\boldsymbol{\psi}$.

(ii) (Asymptotic Normality) $\mathbf{G}_n^{\frac{1}{2}} \left(\hat{\boldsymbol{\psi}}_n^{CL} - \boldsymbol{\psi}^0 \right) \xrightarrow{\mathcal{D}} N(0, \mathbf{I})$ where $\mathbf{G}_n = \mathbf{Q}_n \mathbf{P}_n^{-1} \mathbf{Q}_n$ is the Godambe information matrix (Godambe, 1960). \mathbf{P}_n is the covariance matrix of the gradient $\nabla c\ell_n$ and \mathbf{Q}_n is the negative expected value of the Hessian matrix of $c\ell_n$, where both are evaluated at $\boldsymbol{\psi} = \boldsymbol{\psi}^0$.

Proof. For a composite likelihood, it is sufficient to verify the same regularity conditions as for the usual maximum likelihood estimators (Lindsay, 1988). In the context of expanding domain asymptotics in spatial statistics, the spatial covariance function and its first and second derivatives need to be absolutely summable. From Theorem 3 in Mardia and Marshall (1984), this condition holds for the exponential covariance function that we are using here. (i) and (ii) follow immediately. \square

The main result of this section establishes the consistency and asymptotic normality of the posterior mode, $\hat{\boldsymbol{\psi}}_n^B$.

Proposition 5.3.2. (i) (Posterior consistency) $|\pi_n(\boldsymbol{\psi}) - \pi_n^0(\boldsymbol{\psi})|_{TV} \xrightarrow{\mathcal{P}} 0$ as $n \rightarrow \infty$ where $|\cdot|_{TV}$ is the total variation norm and $\pi_n^0(\boldsymbol{\psi})$ is a normal density with the mean $\boldsymbol{\psi}^0 + \mathbf{Q}_n^{-1} \nabla c\ell_n(\boldsymbol{\psi}^0)$ and the covariance \mathbf{Q}_n^{-1} with $\mathbf{Q}_n^{-1} \rightarrow \mathbf{0}$ as $n \rightarrow \infty$.

(ii) (Asymptotic normality) As $n \rightarrow \infty$, $\mathbf{G}_n^{\frac{1}{2}} \left(\hat{\boldsymbol{\psi}}_n^B - \boldsymbol{\psi}^0 \right) \xrightarrow{\mathcal{D}} N(0, \mathbf{I})$.

Proof. When the maximum composite likelihood estimator $\hat{\boldsymbol{\psi}}_n^{CL}$ is consistent and asymptotically normal, (i) and (ii) follow (Theorems 1 and 2, respectively, in Chernozhukov and Hong (2003)). Hence the result follows directly from Proposition 5.3.1. \square

Application of Gobambe Adjustment. We have several options for adjusting our composite likelihood-based inference. These include (a) direct use of the

asymptotic distribution in Proposition 2. (ii); (b) The ‘open-faced sandwich’ post-hoc adjustment (Shaby, 2014) of MCMC sample from the composite posterior distribution $\pi_n(\boldsymbol{\psi})$; (c) ‘curvature’ adjustment (Ribatet et al., 2012) for our MCMC procedure. We utilize (b) and (c) because these MCMC-based methods can capture the higher-order moments of the posterior distribution, which may be important in finite sample inference.

For any of these methods, it is necessary to evaluate \mathbf{P}_n and \mathbf{Q}_n . See the Appendix B for an example of their analytic computation. Here \mathbf{Q}_n can also be obtained using MCMC runs from the posterior distribution $\pi_n(\boldsymbol{\psi})$ by the asymptotic result in Proposition 2. (i). The matrix operation for computing \mathbf{Q}_n and \mathbf{P}_n is not computationally demanding as it can be done in less than a few minutes using a high-performance single core and needs to be done only once.

We caution that the adjustment procedures here rely on the identifiability of parameters in $\boldsymbol{\psi}$. In order to evaluate \mathbf{P}_n and \mathbf{Q}_n under the correct probability model in (5.6), we need to be able to estimate the true value $\boldsymbol{\psi}^0$ accurately by the posterior mode $\hat{\boldsymbol{\psi}}_n^B$. This may not always hold as there is a trade-off between the discrepancy parameters in $\boldsymbol{\xi}_d$ for finite sample sizes.

The open-faced sandwich adjustment is one approach for adjusting the covariance based on Proposition 5.3.2 (Shaby, 2014). For any MCMC sample of $\boldsymbol{\psi}$ from $\pi_n(\boldsymbol{\psi})$, the open-faced sandwich adjustment is defined by $\tilde{\boldsymbol{\psi}}^{open} = \hat{\boldsymbol{\psi}}_n^B + \mathbf{C}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n^B)$ with $\mathbf{C} = \mathbf{Q}_n^{-1} \mathbf{P}_n^{\frac{1}{2}} \mathbf{Q}_n^{\frac{1}{2}}$. Similar to the curvature adjustment, this approach guarantees that the distribution of the adjusted posterior sample has the same posterior mode and the desired asymptotic covariance \mathbf{G}_n^{-1} . This method can be either embedded in each step of MCMC run, or applied after an entire MCMC run is finished.

Another approach is curvature adjustment (Ribatet et al., 2012), which substitutes $\boldsymbol{\psi}$ in (5.3) with $\tilde{\boldsymbol{\psi}}^{curv} = \hat{\boldsymbol{\psi}}_n^B + \mathbf{D}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n^B)$, where $\boldsymbol{\psi}$ is the posterior mode from (5.3). \mathbf{D} is the matrix that satisfies $\mathbf{D}^T \mathbf{Q}_n \mathbf{D} = \mathbf{Q}_n \mathbf{P}_n^{-1} \mathbf{Q}_n$. This approach ensures that the resulting posterior distribution has the same mode as the original composite likelihood $c\ell_n(\boldsymbol{\psi})$ and the asymptotic covariance \mathbf{G}_n^{-1} , as described in Proposition 2. (ii). The choice for \mathbf{D} is not unique, and Ribatet et al. (2012) suggested using $\mathbf{D} = \mathbf{Q}_n^{\frac{1}{2}} (\mathbf{Q}_n \mathbf{P}_n^{-1} \mathbf{Q}_n)^{\frac{1}{2}}$, where the square roots of the matrices are computed using singular value decomposition. Here we use the open-faced

adjustment; the curvature adjustment approach can also be used but, as shown in (Shaby, 2014), the difference between the two approaches is likely to be minimal.

It is also possible to infer the parameters using the Godambe information matrix. However, Bayesian inference based on the adjustment allows us to incorporate prior information and learn about the shape of the posterior density which might be important for a finite sample. In computer model calibration it is important to use prior information for identifying the discrepancy function; a Bayesian approach is therefore a convenient framework for doing this.

5.4 Application to UVic ESCM Calibration

We demonstrate the application of our approach to a climate model calibration problem. The computer model used here is the University of Victoria Earth system climate model (UVic ESCM) of intermediate complexity (Weaver et al., 2001). The input parameter of interest is climate sensitivity (CS), defined as the equilibrium global mean surface air temperature change due to a doubling of carbon dioxide concentrations in the atmosphere (Andronova et al., 2007; Knutti and Hegerl, 2008). Climate sensitivity is an important model diagnostic and is used as an input to climate projections as well as economic assessments of climate change impacts (see, e.g. Nordhaus and Boyer, 2000; Keller et al., 2004). We use ten UVic ESCM runs with different CS values ($p=10$). Each model run is a spatial pattern of ocean temperature anomaly on a regular 1.8° latitude by 3.6° longitude grid, defined as change between 1955-1964 mean and 2000-2009 mean in degree Celsius times meter ($^\circ\text{C m}$). At each location, the ocean temperature anomaly is vertically integrated from 0 to 2000 m in depth. See Sriver et al. (2012) for a more detailed description of the model runs.

The model output has regions of missing data since it covers only the ocean, and partition of the spatial area needs careful consideration. We partitioned the spatial area using a random tessellation; this is also the approach followed by Eidsvik et al. (2013). We first randomly chose M different centroids out of total n locations and then assigned the spatial locations to different subregions according to the nearest centroid in terms of geodesic distance. When finding the nearest centroid for each point, we only considered the centroids in the same ocean to avoid

assigning locations separated by land to the same block. This random tessellation ensured, on average, that we have more subregions where data points are densely distributed.

Choosing the number of blocks requires balancing two conflicting objectives: using enough number of blocks to make the computation feasible, and ensuring enough sample size for each block for asymptotic convergence. One heuristic approach is to ensure that the maximum distance within each block exceeds the effective range (Eidsvik et al., 2013; Zhang and Zimmerman, 2005), which is the spatial distance at which correlation reduces to 0.05. For the exponential covariance function we use for the discrepancy term in this work, this corresponds to $3\frac{1}{\phi_d}$. In practice, however, it might be difficult to strictly follow this guideline due to the long range dependence for the discrepancy process δ . In such case, one needs to choose the smallest number of blocks that still enables feasible likelihood computation. In a simulated example below, we examine how this affects the inference results.

5.4.1 Simulated Examples

We conducted some perfect model experiments to answer the questions of whether the the posterior density based on the composite likelihood (composite posterior) is similar to the posterior density based on the original likelihood (original posterior), of whether the posterior density with approximated block mean covariance computation (approximated composite posterior) described in (5.4) is close to the true composite posterior, and of how the number of spatial blocks and the magnitude of the discrepancy affect the composite posterior density.

Each experiment followed four key steps: choose one of the parameter settings for model runs as the synthetic truth; leave the corresponding model run out and superimpose a randomly generated error on it to construct a synthetic observation; emulate the computer model using the remaining model runs; calibrate the computer model using the constructed emulator and compare the resulting density with the synthetic truth.

To be able to compute the original posterior density with a reasonable computational effort, we restricted ourselves to a subset of spatial locations consisting of

1000 randomly selected points ($n = 1000$) and assumed separable covariance structure for the spatial field and the computer model parameter space. The synthetic truth for the climate sensitivity used here was 2.153, but choosing other parameter settings gave similar results shown here.

A comparison between the composite posterior densities with 10 blocks and the original posterior densities are shown in Figure 5.1(a) and 5.1(b). We used two different realizations of the model-observation discrepancy. These were generated from a Gaussian process model with exponential covariance (5.5) with $\zeta_d^* = 0.01$, $\kappa_d^* = 160000$, and $\phi_d^* = \frac{1}{690}$ km, where $(\zeta_d^*, \kappa_d^*, \phi_d^*)$ are assumed true values of $(\zeta_d, \kappa_d, \phi_d)$. We also conducted the same comparison for the approximated composite posterior densities (Figure 5.1(c) and 5.1(d)). The posterior densities and the resulting credible intervals from all three approaches are reasonably similar. The composite posterior densities after adjustment are slightly more dispersed than the original posterior due to the information loss caused by blocking, but the modes are quite close to the original ones confirming the consistency result in Proposition 5.3.2 (i).

We also compared the adjusted composite posterior densities with different numbers of blocks to examine the effect of the number of blocks on calibration results (Figure 5.2). The results show that using more than 30 blocks introduces a slight bias for the posterior mode that might be due to the reduced number of data points in each block or the smaller area covered by each block. When using more than 30 blocks, there are several blocks for which the maximum distance is less than the effective range. However, the credible intervals are again reasonably similar to each other and our approach still gives a reasonable answer in this setting. Similarly, we compare the adjusted composite posterior densities based on datasets generated using different assumed sill values, $\kappa_d^* = 40000$, 90000, and 160000 to investigate the effect of magnitude of discrepancy on calibration results (Figure 5.3). As one would expect, the posterior density becomes more dispersed as we increase the value of the sill.

We used informative priors for the statistical parameters, important in reducing the identifiability issues occurring in the calibration based on observational data in Section 5.4.2. We imposed a vague prior for the nugget parameter $\zeta_d \sim IG(2, 0.01(2 + 1))$ and a highly informative prior for the sill parameter

$\kappa_d \sim IG(10000, \kappa_d^*(10000 + 1))$. The sill parameter for the emulator κ_s was given a mildly informative prior with $IG(20, \hat{\kappa}_s(20 + 1))$, $\hat{\kappa}_s$ the MLE of κ_s computed in the emulation stage. The shape parameters for the inverse-Gamma distributions are specified in the way that the prior modes are aligned with certain target values. Note that inference for simulated examples does not suffer from identifiability issues without the informative priors; we use these priors only to be consistent with the calibration based on observational data. For example, changing the shape parameter b_{κ_d} for the inverse-Gamma prior from 10000 to 2 does not change the results in Figures 5.1-5.3.

5.4.2 Calibration Using Observed Ocean Temperature Anomalies

As an illustrative example, we calibrated the climate sensitivity using the observed spatial pattern of ocean temperature anomaly from the data product constructed by Levitus et al. (2012). We interpolated the observational data onto the UVic model grid using a simple bilinear interpolator. By placing the model output and the observational data on the same grid, we take full advantage of the computational gains from the separability assumption. We divided the 5,903 locations ($n = 5903$) into 50 blocks ($M = 50$) using the random tessellation method described above. The covariance matrices for block means were approximated using (5.4) with $m_i = \min(10, n_i)$ for $i = 1, \dots, 50$. The prior specification was the same as in the simulated example with assumed sill (κ_d^*) of 160,000, except that the inverse of the discrepancy range parameter $\frac{1}{\phi_d}$ was restricted to be greater than 800 km to reduce identifiability issues. Figure 5.4 shows the posterior density of climate sensitivity. The length of the MCMC chain was 15,000, and the computing time was about 15 hours (wall time) via parallel computing using 32 high-performance cores for a system with Intel Xeon E5450 Quad-Core 434 at 3.0 GHz. We verified that our MCMC algorithm and chain length were adequate by ensuring that the MCMC standard errors for our parameter estimates (Jones et al., 2006; Flegal et al., 2008) were small enough and by comparing posterior density estimates after various run lengths to see that the results, namely posterior pdfs, had stabilized.

Unlike simulated examples described above, the calibration results with and

without the adjustment described in Section 5.3.2 were clearly different due to the reduced block sizes. This indicates that the adjustment using the Godambe information is an important step to account for the information loss caused by blocking, especially when using many smaller blocks is unavoidable due to the large amount of data. In our particular application blocking causes underestimation of the variance of the parameter θ . Underestimation of the variances of important input parameters can lead to an incorrect analysis in follow-up studies based on the calibration results. For example, underestimating the uncertainty in the important climate parameters such as the climate sensitivity can result in incorrect risk assessment in climate change impact studies.

5.5 Discussion

5.5.1 Caveats

While our approach is helpful in mitigating computational issues for various calibration problems, there is still more work to be done to make the computation more efficient. As n continues to get large the number of spatial locations in each block may become excessively large and evaluation of composite likelihood may not be computationally tractable. One can consider increasing the number of blocks until the computation becomes feasible, but then the convergence of the posterior modes may be very slow due to too small block sizes (Cox and Reid, 2004; Varin, 2008). Another perhaps simpler approach is to use a composite likelihood framework that does not involve blocks though this may involve the need for analytical work to establish posterior propriety.

Another possible issue is related to the use of a Gaussian emulator in place of the true computer model in computing \mathbf{P}_n and \mathbf{Q}_n . Using a Gaussian process emulator, we approximate not only the true computer model itself, but also its first and second derivatives. In our particular example, this did not cause any problem due to very regular behavior of the computer model output with respect to the input parameters. However, this may not be true in general and therefore \mathbf{P}_n and \mathbf{Q}_n calculations may be inaccurate.

The asymptotic independence between input parameters and discrepancy pa-

rameters does not usually hold in a finite sample. It is well known that calibration models usually suffer from identifiability issues (Wynn, 2001). One way to avoid the issues is to impose discrepancy prior information on the discrepancy term (Arendt et al., 2012) as we did in Section 5.4.

The scientific result shown in 5.4.2 requires some caution in its interpretation. First, besides climate sensitivity, climate system response to changes in radiatively active gases in the atmosphere also depends on the magnitude of the radiative effects of these gases (“radiative forcing”), and on the vertical mixing of heat into the deep ocean (Hansen et al., 1985; Knutti et al., 2002; Schmittner et al., 2009; Urban and Keller, 2010). The parameters controlling the forcing and the vertical mixing were kept fixed, respectively, at 0.1 and 1.5 in the model runs we use. Including these additional uncertainties is expected to make the posterior density of CS more dispersed. The example serves as a demonstration of computational feasibility of our approach when applied to high-dimensional spatial datasets rather than providing an improved estimate of CS. Second, the variability of the posterior density is sensitive to the prior information for the discrepancy term; this is a common problem for many calibration problems, as discussed earlier.

5.5.2 Summary and Future Direction

This work is, to our knowledge, the first application of composite likelihood to the computer model calibration problem. Our composite likelihood approach enables computationally efficient inference in computer model calibration using high-dimensional spatial data. We proved consistency and asymptotic normality of our posterior estimates and established covariance adjustment for posterior density based on them. The adjustment can be easily integrated into common MCMC algorithms such as the Metropolis-Hastings algorithm. The block composite likelihood used here yields a valid probability model, and therefore no additional verification for the propriety of the posterior distribution is necessary.

An attractive benefit of this general framework is that it is relatively easy, in principle, to extend the approach to a more complicated and easy-to-interpret covariance model. For example, by allowing covariance parameters to vary across the different spatial blocks, our approach can introduce non-stationarity in the

spatial processes of model output and observational data.

So far one underlying assumption in our study is that computer model output and observational data can be adequately modeled by multivariate normal distributions given by Gaussian processes. However, this is not the case for many important applications such as ice sheet model calibration where the variables in model output and observational data are highly non-Gaussian. In the next chapter our objective is to develop a calibration framework for non-Gaussian spatial data.

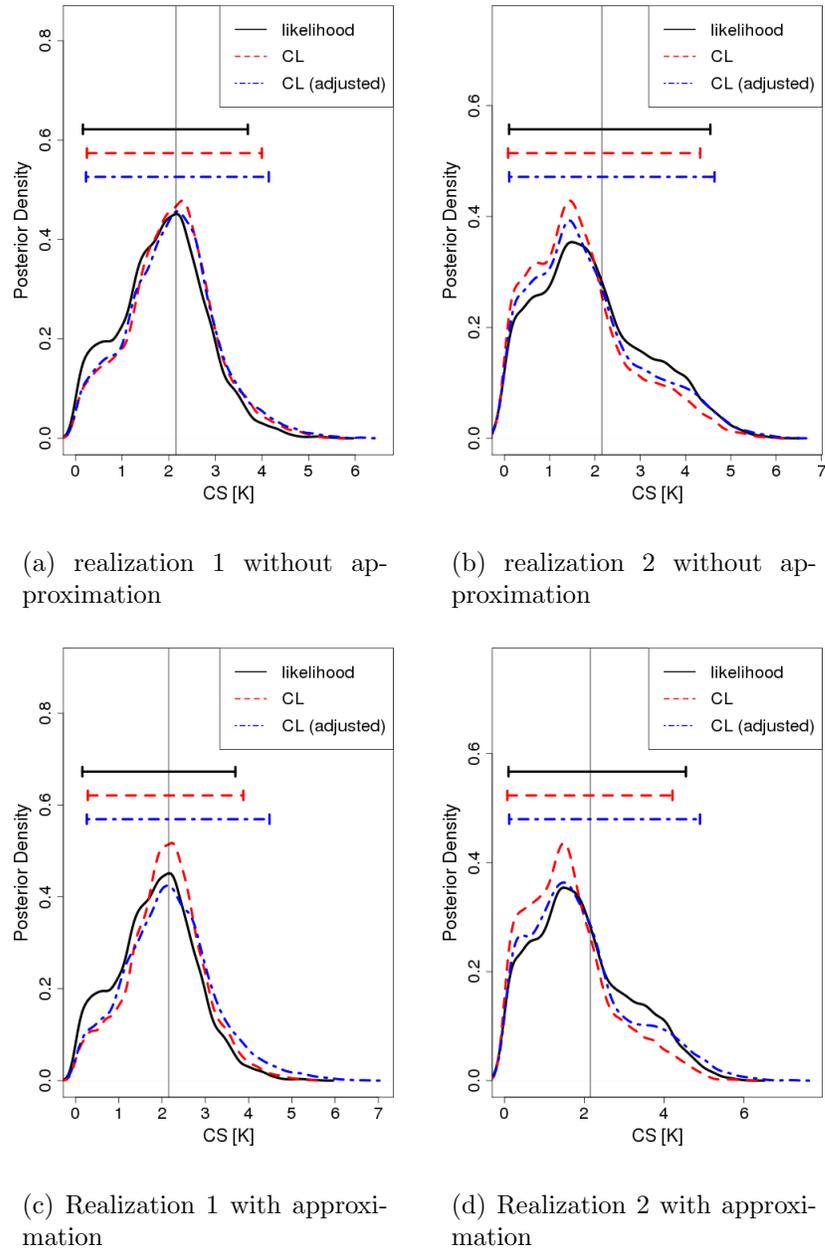


Figure 5.1: Comparison between calibration results using i) the original likelihood without blocking (solid black curves), ii) the block composite likelihood without the variance adjustment (dashed red line), and iii) the block composite likelihood with the variance adjustment (dashed-dotted blue line). The vertical lines represent the assumed true value for our simulation, and the horizontal bars above show the 95% credible intervals. The results shown here are based on two different realizations (two left panels for Realization 1 and two right panels for Realization 2) from the same Gaussian process model. The posterior densities with the approximation for the block means (two lower panels) are reasonably close to the densities without the approximation (two upper panels) when the variance adjustment is applied.

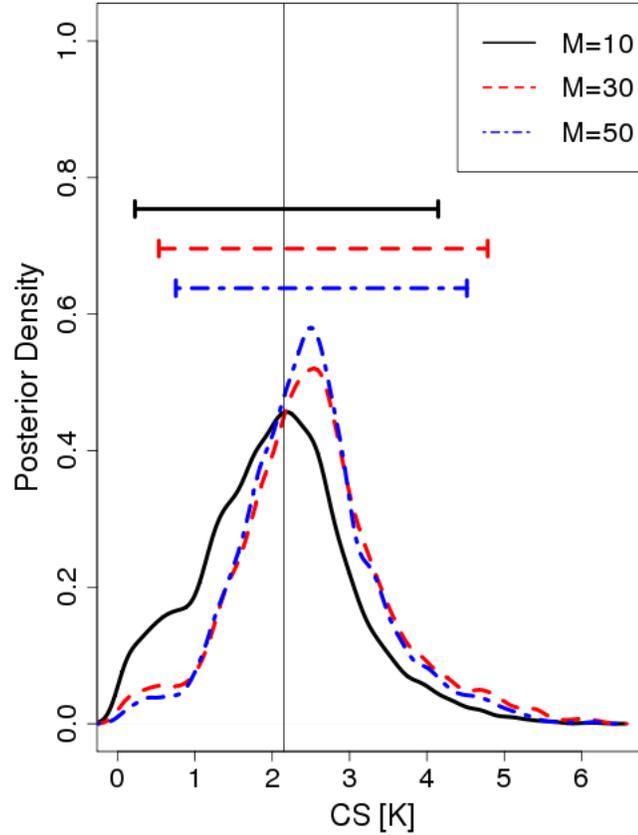


Figure 5.2: Comparison of posterior densities between three simulated examples with different block numbers: $M = 10$ (solid black curve), $M = 30$ (dashed red curve), and $M = 50$ (dotted-dashed blue curve). The vertical line is the assumed true value for our simulated example and the horizontal bars above are 95% credible intervals. Posterior modes based on 30 and 50 blocks show slight biases, but the width of intervals does not show notable differences.

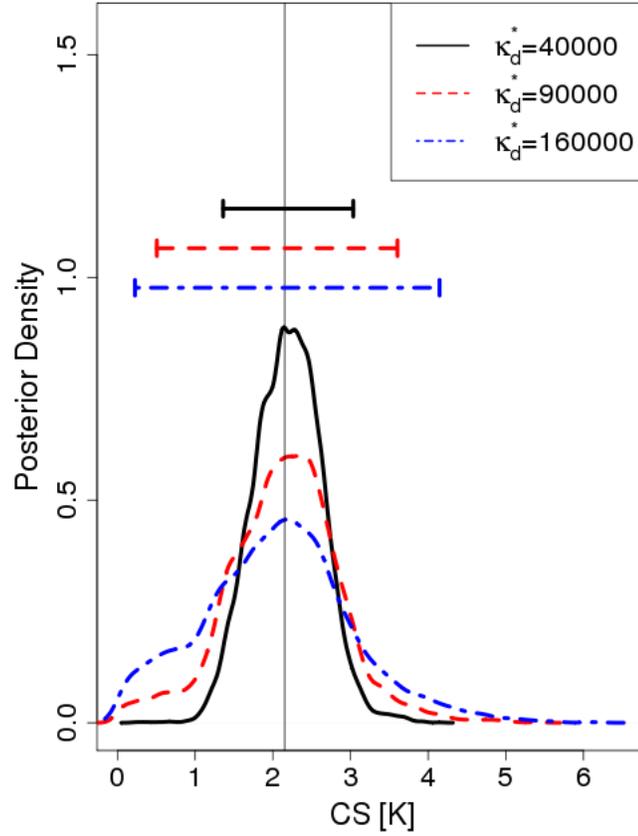


Figure 5.3: Comparison of posterior densities between three simulated examples with different assumed magnitudes of the discrepancies: $\kappa_d^* = 40000$ (solid black curve), $\kappa_d^* = 90000$ (dashed red curve), and $\kappa_d^* = 160000$ (dotted-dashed blue curve). The vertical line indicates the assumed true value, and the horizontal bars above show the 95% credible intervals. As the discrepancy grows, the densities become more dispersed but the posterior modes stay similar.

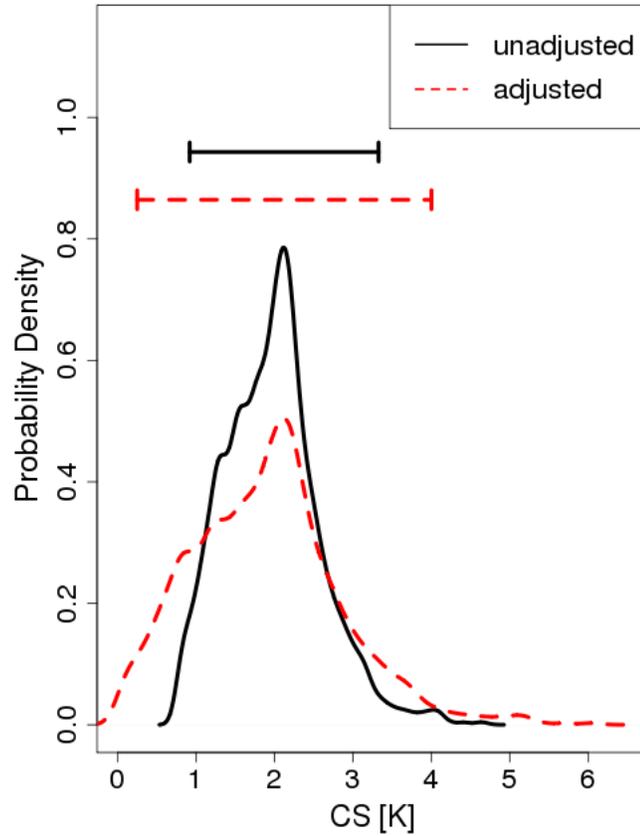


Figure 5.4: Posterior densities of the climate sensitivity calibrated based on the observational data from Levitus et al. (2012) using our composite likelihood approach. The adjusted posterior density (solid black curve) is notably more dispersed than the unadjusted one (dashed black curve), and the corresponding 95% credible intervals (horizontal bars above) for the adjusted posterior density is also much wider than the one for the unadjusted density.

Ice Sheet Model Calibration Using High-Dimensional Binary Spatial Data*

In the previous two chapters we study the calibration frameworks for datasets that are well modeled by Gaussian processes. For many computer models including ice sheet models, however, outputs are often non-Gaussian and the current approaches including ours do not provide suitable probability models for such data sets. In this section we develop a new calibration method for non-Gaussian spatial data. Our method generalizes the approach we presented in Chapter 4 to non-Gaussian spatial data that may be modeled (marginally) by distributions in the one-parametric exponential family. Similar to the calibration framework for Gaussian processes, this approach faces computational and inferential challenges when the model output and observational data are high-dimensional spatial data. Care also needs to be taken when specifying an appropriate error structure, particular for data-model discrepancies. To mitigate these challenges we propose and study a computationally efficient calibration framework based on dimension reduction using principal component analysis for the one parameter exponential family. We demonstrate the performance of our approach by applying it to a simple numerical example and a Greenland ice sheet model calibration problem.

*This chapter is based on a manuscript in preparation with Murali Haran, Patrick Applegate, and Klaus Keller.

6.1 Motivating Problem from Ice Sheet Model

There is increasing evidence that climate change due to human activity affects the Greenland ice sheet (Alexander et al., 2013; Alley et al., 2005). Characterizing and predicting the behavior of the Greenland ice sheet is important in climate change impacts and risk management due to its potentially significant contribution to future sea level rise (Bamber et al., 2001; Bamber and Aspinall, 2013; Lemke et al., 2007). Generating projections for future ice volume changes using ice sheet models is one of the main avenues of approach to this problem. Ice sheet models include input parameters that control modeled ice sheet behavior, including the evolution of ice sheet volume over time and the spatial distribution of ice.

Applegate et al. (2012) assesses the effect of parameter configuration on projections from the model SICOPOLIS (Greve, 1997; Greve et al., 2007) and shows that the future ice volume change projections have large variation between input parameter settings. Their findings indicate that parametric uncertainty is an important source of the projection uncertainty, and therefore characterizing and, when possible, reducing parametric uncertainty is a key to produce ice sheet projections that can properly inform climate risk analysis.

Model calibration using observational data is a way to characterize the parametric uncertainty. Applegate et al. (2012) uses a simple filtering approach selecting parameter settings that yield ice volumes within 10% of the observed modern ice volume. Although this simple method leads to a substantial parametric uncertainty reduction, the result still leaves room for improvement; (i) the approach reduces uncertainty for only one parameter out of the five and therefore further uncertainty reduction may be possible, and (ii) the results do not allow for probabilistic interpretation and hence more rigorous approach based on a statistical model can provide a better interpretability.

Gladstone et al. (2012) uses the confidence set approach, which chooses input parameter settings that yield output values which fall in the 95% ‘confidence set’. Their approach allows valid probabilistic interpretation but the model-observation discrepancy and the observational errors are specified only by expert judgment. Moreover, the ice sheet model needs to be computationally very efficient so that it can be run for at least thousands of parameter settings. McNeall et al. (2013) uses

a Gaussian process emulator for three model output summaries, the overall ice volume, the maximum ice height, and the area covered by ice. Their approach can be applicable to computationally expensive ice sheet models due to the Gaussian process emulator. However, instead of utilizing the probability model given by the emulator they rely on an “implausibility measure”, which is the model output mean standardized using the observational data. Both approaches enable probabilistic interpretation for ice sheet model projection to some extent, but the information provided by their results is still limited because they only provide approximate confidence limits for the input parameters and the correspond projections.

The approach in Chang et al. (2013a) improves upon these two approaches by conducting a full probabilistic calibration based on the probability model for their Gaussian process emulator and using spatially resolved information from average ice thickness profile for latitude (see Section 6.2 for more details). The probabilistic calibration approach using a Gaussian process emulator (Kennedy and O’Hagan, 2001) has been successfully applied to climate model calibration problems (Sansó and Forest, 2009; Bhat et al., 2012; Chang et al., 2014) and has provided a statistically sound way to generate future climate projections. However, the current approaches that fall into this category are not applicable to data in the form of non-Gaussian processes; non-Gaussian spatial data are important sources of spatial information on the Greenland ice sheet such as the extent of ice covered area in modern- and paleo-observations (cf. Alley et al., 2005). Chang et al. (2013a) circumvents this problem by using datasets obtained by aggregating the 2-dimensional spatial pattern of ice thickness into 1-dimensional profile and hence is silent on using datasets that are in the form of non-Gaussian spatial processes.

In this chapter we propose a calibration approach for binary spatial data. To our knowledge this is the first attempt to propose a rigorous calibration framework for non-Gaussian spatial data. The procedure is similar to the calibration framework for Gaussian data described in Chapter 2; we construct an emulator that interpolates between binary model outcomes at different input parameter settings and use the probability model given by the emulator to set up a calibration model. The approach enables utilization of important sources of information that cannot be adequately modelled by the current calibration approaches such as the extent of ice covered areas. Since binary spatial patterns from ice sheet model

output and observations are typically high-dimensional, calibration problem using these datasets poses computational and inferential challenges. To mitigate these challenges we propose a calibration framework based on dimension reduction using logistic principal component analysis (Collins et al., 2002; Lee et al., 2010). This framework is based on dimension reduction for natural parameters for the generalized linear model, and therefore it is in principle applicable to any calibration problem where the data may have been plausibly generated from a one parametric exponential family distribution.

The remainder of this chapter is organized as follows. In Section 6.2 we briefly review the ice sheet model calibration approach in Chang et al. (2013a). In Section 6.3 we introduce the basic framework for calibration using non-Gaussian spatial data and explain computational challenges posed by high-dimensional spatial data. We formulate a reduced-dimensional approach to mitigate these challenges in Section 6.4. We describe how to compute the logistic principal components (PCs) for binary computer model output and formulate an emulation framework based on them, and then set up a calibration model that relies on the reduced-dimensional emulator. Section 6.5 describes the implementation details for our experiments on a simulated example and a Greenland ice sheet model and presents the results from those experiments. We conclude the chapter by discussing future directions and caveats in Section 6.6.

6.2 Ice Sheet Model Calibration Using Ice Thickness Profile

Chang et al. (2013a) calibrates an ice sheet model by working with aggregated ice sheet data. Aggregation has the effect of turning the ice sheet data into data more amenable to Gaussian process modeling. The calibration method used in Chang et al. (2013a) is based on dimension reduction using principal components and kernel convolution (Higdon et al., 2008; Chang et al., 2014). Based on the average ice thickness profile along latitudes, this approach performs full probabilistic calibration based on a Gaussian process emulator for the principal components of the model output. This approach improves the previous results in two

aspects: First, this approach yields the posterior distribution of the input parameters, which provides much richer information such as interaction between input parameters and yields predictive densities of the corresponding projections. Second, this approach uses spatially resolved information rather than a few summary statistics and therefore results in smaller variations for the input parameters and the projections. However, in this approach the 2D ice thickness patterns are aggregated into 1D latitude profiles to avoid dealing with zero-inflated spatial data, which poses nontrivial computational and inferential challenges.

6.3 Computer Model Emulation and Calibration Using Binary Spatial Data

We first introduce a basic framework for computer model emulation and calibration using binary spatial data and explain the inferential and computational challenges that motivate our reduced-dimension approach described in the next section. We focus on the binary data case for ease of exposition, but the framework applies more generally to the one-parametric exponential family. The basic calibration framework for Gaussian spatial data (see e.g. Bhat et al., 2010, 2012; Chang et al., 2014, 2013b) is also a special case of this approach. Here we adapt the two stage approach (Bayarri et al., 2007; Bhat et al., 2012; Chang et al., 2014, 2013b) where we approximate the computer model by interpolating the output first (emulation step) and then conduct calibration using the emulator (calibration step). Separating the emulation step from the calibration step makes it easier to diagnose the performance of the emulator and improves the identifiability of the parameters in the calibration model (see Bhat et al., 2010, 2012, for more discussion).

We use the following notation henceforth. Let $Y(\boldsymbol{\theta}, \mathbf{s})$ be the binary computer model output at a parameter setting $\boldsymbol{\theta} \in \Theta$ and a spatial location $\mathbf{s} \in \mathcal{S}$ where $\Theta \in R^d$ is the parameter space and \mathcal{S} is the spatial field that covers the geographical area of interest. The integer d represents the number of input parameters that the computer model has. We assume that we have obtained model runs at p design points $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p \in \Theta$ and each model run has n spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathcal{S}$. We let \mathbf{Y} be the a $p \times n$ matrix of the existing computer model output where its

(i, j) th element $Y(\boldsymbol{\theta}_i, \mathbf{s}_j)$ is the model output at the i th parameter setting $\boldsymbol{\theta}_i$ and the j th spatial location \mathbf{s}_j . Similarly we denote \mathbf{Z} be the n -dimensional observational data where its j th element $Z(\mathbf{s}_j)$ is the observation at the j th spatial location \mathbf{s}_j .

6.3.1 Computer Model Emulation Using Binary Spatial Data

Let $\boldsymbol{\Gamma}$ be the $p \times n$ matrix of natural parameters for model output, where its (i, j) th element $\gamma_{ij}^Y = \log \frac{p_{ij}}{1-p_{ij}}$ is the logit for the i th parameter setting at the j th spatial location and $p_{ij} = P(Y(\boldsymbol{\theta}_i, \mathbf{s}_j) = 1)$. The probability model for $Y(\boldsymbol{\theta}_i, \mathbf{s}_j)$ is given by

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

and therefore the log likelihood function for $\boldsymbol{\Gamma}$ given the model output \mathbf{Y} can be rewritten as

$$\ell(\boldsymbol{\Gamma}; \mathbf{Y}) = \sum_{i=1}^p \sum_{j=1}^n P(Y(\boldsymbol{\theta}_i, \mathbf{s}_j) = y_{ij}) = \sum_{i=1}^p \sum_{j=1}^n (1 + \exp(-(2y_{ij} - 1)\gamma_{ij}))^{-1}. \quad (6.1)$$

To approximate the computer model we interpolate the natural parameters at different input parameter settings using a Gaussian process (cf. Sacks et al., 1989) with zero mean and the following covariance function:

$$\text{Cov}(\gamma_{ij}, \gamma_{kl}) = C(\boldsymbol{\theta}_i, \boldsymbol{\theta}_k, \mathbf{s}_j, \mathbf{s}_l; \boldsymbol{\xi}_y),$$

where $\boldsymbol{\xi}_y$ is a vector of the covariance parameters. Note that estimating the natural parameters $\{\gamma_{ij}, i = 1, \dots, p, j = 1, \dots, n\}$ is not possible by simply maximizing (6.1) since this is an ill-posed problem without any constraints on them (see Section 6.3.3 for details). If estimates of the natural parameters are available, we can fit the Gaussian process to those natural parameters by finding MLE of $\boldsymbol{\xi}_y$ and use the fitted process to predict the the process of natural parameters for any untried

parameter setting $\boldsymbol{\theta}$ by the standard kriging approach (Cressie, 1993; Stein, 1999). We call the fitted Gaussian process an “emulator” and the resulting predictive process an “emulator process”.

6.3.2 Computer Model Calibration Using Binary Spatial Data

We set up a calibration model that links the input parameter to the observational data using the emulator constructed in the previous section while considering the systematic discrepancy. Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^T$ be the n -dimensional vector of natural parameters for the observational data, then our calibration model is given by

$$\boldsymbol{\lambda} = \boldsymbol{\eta}(\boldsymbol{\theta}^*, \mathbf{Y}) + \boldsymbol{\delta},$$

where $\boldsymbol{\delta}$ is an n -dimensional vector for the model-observation discrepancy. For $\boldsymbol{\delta}$ we use an n -dimensional Gaussian process with a vector of covariance parameters $\boldsymbol{\xi}_d$ that is independent of the emulator process $\boldsymbol{\eta}(\boldsymbol{\theta}^*, \mathbf{Y})$. This corresponds to the following probability model for each observation given by the standard logistic regression framework:

$$\begin{aligned} P(Z(\mathbf{s}_i) = z_i) &= \left(\frac{\exp(\lambda_i)}{1 + \exp(\lambda_i)} \right)^{z_i} \left(\frac{1}{1 + \exp(\lambda_i)} \right)^{1-z_i} \\ &= (1 + \exp(-(2z_i - 1)\lambda_i))^{-1} \end{aligned}$$

where λ_i is the i th element of $\boldsymbol{\lambda}$. With a standard prior specification for the input parameters $\boldsymbol{\theta}^*$ and the discrepancy parameters in $\boldsymbol{\xi}_d$, we can define the posterior density for those parameters and make an inference using MCMC based on the posterior density.

6.3.3 Computational Challenges

The framework described above faces computational and inferential challenges when the number of spatial locations n is large. The emulation step described in Section 6.3.1 requires to estimate the np different natural parameters. This is an ill-posed problem since the number of data points in the model output is also np .

Finding a set of constraints that leads to an estimable number of parameters while still allowing enough flexibility for the resulting process is not straightforward. Although estimates of the natural parameters are available, emulation for those estimates still poses significant computational challenges. The emulation involves numerically finding the maximum likelihood estimate of $\boldsymbol{\xi}_y$, requiring repeated evaluation of a computationally infeasible likelihood function; each likelihood evaluation here involves Cholesky decomposition of the $np \times np$ covariance matrix, whose computation scales as $\mathcal{O}(n^3p^3)$. For the example in Section 6.5 this translates to computational cost of $\frac{1}{3} \times 99^3 \times 22,082^3 = 3.48 \times 10^{18}$ flops, which in turn corresponds to at least hundreds of thousands of computing time for a 3.0 Ghz single core. The calibration step is subject to similar challenges. Estimating the input parameters in $\boldsymbol{\theta}^*$ and the discrepancy parameters in $\boldsymbol{\xi}_d$ requires integrating out the n -dimensional discrepancy process $\boldsymbol{\delta}$, and it makes obtaining a well-mixed MCMC chain very difficult even for a moderate size of n . Moreover, each likelihood evaluation involves Cholesky decomposition of the $n \times n$ covariance matrix and its computational cost scales as $\mathcal{O}(n^3)$. This translates to computational cost of $\frac{1}{3} \times 22,082^3 = 3.59 \times 10^{12}$ flops in the example in Section 6.5, which corresponds to half an hour of computing time for a 3.0 Ghz single core.

6.4 Reduced-Dimension Approach for Binary Spatial Data

We propose a reduced-dimension approach to overcome the computational challenges discussed in the previous section. In the emulation step we build an emulator for the logistic principal components of the model output to solve the inferential issues in estimating np different natural parameters and the computational issues for dealing with the $np \times np$ covariance matrix. In the calibration step we use a basis representation approach to model the discrepancy process, which removes the need to integrate out the n -dimensional discrepancy process and compute the Cholesky decomposition of the $n \times n$ covariance matrix.

6.4.1 Dimension Reduction for Binary Model Output

In this subsection we describe the logistic principal component analysis for binary model output in detail by closely following Lee et al. (2010). For dimension reduction we assume that $\mathbf{\Gamma}$ can be written as

$$\mathbf{\Gamma} = \mathbf{1}_p \otimes \boldsymbol{\mu}^T + \mathbf{W}\mathbf{K}_y^T, \quad (6.2)$$

where \mathbf{K}_y is an $n \times J_y$ orthogonal basis matrix, \mathbf{W} is the $p \times J_y$ principal component matrix with (i, j) th element $w_j(\boldsymbol{\theta}_i)$, and $\boldsymbol{\mu}$ is the $n \times 1$ mean vector. We rewrite the likelihood in (6.1) as

$$\ell(\boldsymbol{\mu}, \mathbf{K}_y, \mathbf{W}; \mathbf{Y}) = \sum_{i=1}^p \sum_{j=1}^{J_y} g(y_{ij}^*(\mu_j + \mathbf{w}_i \mathbf{k}_{y,j}^T)), \quad (6.3)$$

where $g(x) = (1 + \exp(-x))^{-1}$, μ_j is the j th element of $\boldsymbol{\mu}$, $y_{ij}^* = 2y_{ij} - 1$, \mathbf{w}_i is the i th row of \mathbf{W} , and $\mathbf{k}_{y,j}$ is the j th column of \mathbf{K}_y .

Following Lee et al. (2010), we estimate the matrix \mathbf{W} and \mathbf{K}_y by maximizing the likelihood in (6.3) using the majorization-minimization (MM) algorithm (Lange et al., 2000; Hunter and Lange, 2004). For each iteration of the MM algorithm we minimize the majorizing function instead of the original likelihood function. The majorizing function is a function that is (i) of the same value as the original likelihood at the current value of the iteration, (ii) greater than the original likelihood for all other values of the parameters, and (iii) in an easier form to maximize than the original likelihood, such as a quadratic function. As a first step, we note that

$$-\log g(x) \leq -\log g(y) - (1 - g(y))(x - y) + \frac{1}{8}(x - y)^2. \quad (6.4)$$

for a given value y . The equality holds if $x = y$. We can complete the square by adding a term that does not depend on y to rewrite the right hand side as follows;

$$-\log g(x) \leq -\log g(y) + \frac{1}{8}(x - y - 4(1 - g(y)))^2. \quad (6.5)$$

By the fact that y_{ij}^* can take only the value of either +1 or -1, plugging in $y_{ij}^* \gamma_{ij}$

and $y_{ij}^* \gamma_{ij}^{(m)}$ into x and y respectively yields

$$-\log g(y_{ij}^* \gamma_{ij}) \leq -\log g(y_{ij}^* \gamma_{ij}^{(m)}) + \frac{1}{8} \left(\gamma_{ij} - x_{ij}^{(m)} \right)^2, \quad (6.6)$$

where γ_{ij} is the (i, j) th value of $\mathbf{\Gamma}$, $\gamma_{ij}^{(m)}$ is the value of γ_{ij} at the m th iteration, and $x_{ij}^{(m)} = \gamma_{ij}^{(m)} + 4y_{ij}^* (1 - \pi(y_{ij}^* \gamma_{ij}^{(m)}))$. Note that (6.2) implies that $\gamma_{ij} = \mu_j + \mathbf{w}_i \mathbf{k}_{y,j}^T$ and $\gamma_{ij}^{(m)} = \mu_j^{(m)} + \mathbf{w}_i^{(m)} \mathbf{k}_{y,j}^{(m)T}$ where $\mu_j^{(m)}$, $\mathbf{k}_{y,j}^{(m)}$, and $\mathbf{w}_i^{(m)}$ are the values of μ_j , $\mathbf{k}_{y,j}$, and \mathbf{w}_i at the m th iteration respective. Since $-\log g(y_{ij}^* \gamma_{ij}^{(m)})$ does not depend of γ_{ij} , the majoring function for the negative likelihood function in (6.3) is given by

$$\sum_{i=1}^n \sum_{j=1}^p \left(x_{ij}^{(m)} - \mu_j - \mathbf{w}_i \mathbf{k}_{y,j}^T \right)^2.$$

For each iteration the minimum of the majorization function occurs at

$$\mu_j = \frac{1}{p} \sum_{i=1}^n \left(x_{ij}^{(m)} - \mathbf{w}_i \mathbf{k}_{y,j}^T \right), \quad (6.7)$$

$$\mathbf{w}_i = (\mathbf{K}_y^T \mathbf{K}_y)^{-1} \mathbf{K}_y^T (\mathbf{x}_i^{(m)T} - \boldsymbol{\mu}), \quad (6.8)$$

$$\mathbf{k}_j = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \left(\mathbf{x}_{\cdot j}^{(m)} - \mathbf{1}_p \mu_j \right), \quad (6.9)$$

where $\mathbf{x}_i^{(m)}$ and $\mathbf{x}_{\cdot j}^{(m)}$ are the i th row and the j th column of $\mathbf{X}^{(m)}$ respectively and $\mathbf{X}^{(m)}$ is a $p \times n$ matrix having $x_{ij}^{(m)}$ as its (i, j) th element. Cycling through (6.7) - (6.9) results in a local maxima of the likelihood function in (6.3). We summarize the algorithm as follows.

1. Initialization: Set $m = 1$ and choose starting points $\boldsymbol{\mu}^{(1)} = (\mu_1^{(1)}, \dots, \mu_p^{(1)})^T$, $\mathbf{W}^{(1)} = (\mathbf{w}_1^{(1)}, \dots, \mathbf{w}_p^{(1)})^T$ and $\mathbf{K}_y^{(1)} = (\mathbf{k}_1^{(1)}, \dots, \mathbf{k}_{j_y}^{(1)})$.
2. Compute $\gamma_{ij}^{(m)} = \mu_j^{(m)} + \mathbf{w}_i^{(m)} \mathbf{k}_{y,j}^{(m)T}$, $x_{ij}^{(m)} = \gamma_{ij}^{(m)} + 4y_{ij}^* (1 - \pi(y_{ij}^* \gamma_{ij}^{(m)}))$ and let $\mathbf{X}^{(m)} = \{x_{ij}^{(m)}\}$.
3. Update $\boldsymbol{\mu}^{(m+1)} = \frac{1}{p} \left(\mathbf{X}^{(m)} - \mathbf{W}^{(m)} \mathbf{K}_y^{(m)T} \right)^T \mathbf{1}_p$.
4. Compute $\mathbf{W}^* = (\mathbf{X}^{(m)} - \mathbf{1}_p \otimes \boldsymbol{\mu}^{(m+1)T})^T \mathbf{K}_y^{(m)} \left(\mathbf{K}_y^{(m)T} \mathbf{K}_y^{(m)} \right)^{-1}$ and find its QR decomposition $\mathbf{W}^* = \mathbf{Q}\mathbf{R}$. Update $\mathbf{W}^{(m+1)} = \mathbf{Q}$.

5. Update $\mathbf{K}_y^{(m+1)} = (\mathbf{X}^{(m)} - \mathbf{1}_p \otimes \boldsymbol{\mu}^{(m+1)T})^T \mathbf{W}^{(m+1)}$.
6. Set $m + 1$ and repeat from 2 until convergence.

6.4.2 Emulation Using Logistic PCs

As in Chapter 4 we fit a Gaussian process for each column of \mathbf{W} separately to construct a J_y -dimensional emulator process. By noting that the i th row of \mathbf{W} corresponds to the i th parameter setting $\boldsymbol{\theta}_i$ we rewrite the (i, j) th of \mathbf{W} as $w_{ij} = w_j(\boldsymbol{\theta}_i)$ and set up a Gaussian process for each j th column with the following squared exponential covariance function:

$$\text{Cov}(w_j(\boldsymbol{\theta}_k), w_j(\boldsymbol{\theta}_l)) = \kappa_{y,j} \exp\left(-\sum_{i=1}^5 \frac{|\theta_{ik} - \theta_{il}|^2}{\phi_{y,ij}^2}\right) + \zeta_{y,j} 1(\boldsymbol{\theta}_k = \boldsymbol{\theta}_l),$$

with the partial sill $\kappa_{y,j}$, the range parameters $\phi_{y,ij}$, the nugget parameter $\zeta_{y,j}$, and the i th element of $\boldsymbol{\theta}_k$, θ_{ik} . Using the squared exponential covariance function may result in an overly regularly behaving process (Stein, 1999), but in our case the model with squared exponential covariance shows a better performance than the model with exponential covariance in cross-validation (cf. Chang et al., 2014). We estimate these parameters via MLE. The computational cost for each likelihood evaluation is $\frac{1}{3}p^3$ flops, which is $\frac{1}{3}100^3 \approx 3.33 \times 10^5$ flops in our application.

The J_y -dimensional emulator process $\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{W})$ for the logistic PCs at any untried parameter setting $\boldsymbol{\theta}$ is given as the standard kriging framework (Cressie, 1993). Then the n -dimensional emulator process for the natural parameters at the untried parameter setting $\boldsymbol{\theta}$ is

$$\boldsymbol{\gamma}(\boldsymbol{\theta}) = \boldsymbol{\mu} + \mathbf{K}_y \boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{W}), \quad (6.10)$$

and the corresponding probability for the predicted model output at each location to be one is

$$p_j(\boldsymbol{\theta}) = P(Y(\boldsymbol{\theta}, \mathbf{s}_j) = 1) = \frac{\exp(\gamma_j(\boldsymbol{\theta}))}{1 + \exp(\gamma_j(\boldsymbol{\theta}))},$$

where $\gamma_j(\boldsymbol{\theta})$ is the j th element of $\boldsymbol{\gamma}(\boldsymbol{\theta})$ in (6.10). To translate the probabilities $p_1(\boldsymbol{\theta}), \dots, p_n(\boldsymbol{\theta})$ into predictions for the binary outcome we simply dichotomize

them using 0.5 as a cut-off.

6.4.3 Calibration Using Logistic PCs

Once the emulator for the logistic PCs is available we set up the following calibration model for the n -dimensional vector of the natural parameter $\boldsymbol{\lambda}$ for observational data using the emulator $\mathbf{K}_y\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{W})$:

$$\boldsymbol{\lambda} = \boldsymbol{\mu} + \mathbf{K}_y\boldsymbol{\eta}(\boldsymbol{\theta}^*, \mathbf{W}) + \boldsymbol{\delta},$$

where $\boldsymbol{\theta}^*$ is the best fit value for the observational data and $\boldsymbol{\delta}$ is the discrepancy term that represents structural error between the model output and the observational data.

To get around the challenges in integrating out $\boldsymbol{\delta}$ described above we use a basis representation for the discrepancy term such that

$$\boldsymbol{\delta} = \mathbf{K}_d\mathbf{v},$$

with the $n \times J_d$ basis matrix \mathbf{K}_d and the J_d -dimensional random coefficient vector $\mathbf{v} \sim N(\mathbf{0}, \sigma_d^2\mathbf{I}_{J_d})$. Finding the basis \mathbf{K}_d that can precisely represent the discrepancy process with a small number of basis vectors is important for computational tractability. To this end we use the following procedure that is inspired by the definition of the discrepancy term:

1. For each parameter setting $\boldsymbol{\theta}_i$ we compute the proportion of match with observational data by $q_i = \sum_{j=1}^n I(Z(\mathbf{s}_j) = Y(\boldsymbol{\theta}_i, \mathbf{s}_j))/n$ where $I(\cdot)$ is the indicator function. Find the ‘best’ model runs with $q_i > c$ for a constant $0 < c < 1$ and let \mathcal{B} be the set of indexes for the selected parameter settings.
2. For each location \mathbf{s}_j compute the proportion of mismatch between the model output and the observational data for the selected parameter settings in 1 by $r_j = \sum_{i \in \mathcal{B}} I(Z(\mathbf{s}_j) \neq Y(\boldsymbol{\theta}_i, \mathbf{s}_j))/p$.
3. Set $J_d = 1$ and hence \mathbf{K}_d is reduced to a n -dimensional vector \mathbf{k}_d . Define the j th element of \mathbf{k}_d as $I(0.5 < r_j)$. The random coefficient vector is also reduced to a univariate random variable $v \sim N(0, \sigma_d^2)$.

The main goal of the above procedure is to capture a common discrepancy pattern that is persistent across all parameter settings which yield reasonably close model output to the observational data. Note that the cut-off value c used to select the best model runs in Step 1 needs to be determined a priori. In the examples described in Section 6.5.1 and 6.5.2 this procedure with $c = 0.6$ can identify the spatial locations where the binary outcomes in the observational data has altered from 0 to 1 or 1 to 0 by the superimposed discrepancy process (See Figure 6.1) and therefore we use this value of c henceforth.

6.5 Implementation Details and Results

We demonstrate the performance of our calibration approach through an example of (i) a simple simulated example and (ii) a real ice sheet model. For each case our objective is to understand the properties of our calibration approach through the perfect model experiment where we calibrate the ice sheet model using synthetic observations. To obtain the synthetic observations we select one model run as the synthetic truth and contaminate it using a spatially correlated process to simulate presence of systematic model-observation discrepancy. Our objective is then to recover the input parameter setting for the synthetic truth using our emulation and calibration method.

6.5.1 Simulated Example

Before we proceed to a problem of calibrating a complex ice sheet model we run a perfect model experiment using a simple numerical example. Unlike the real ice sheet model calibration problem below, this simple example does not have complicating issues such as identifiability of input parameters or sparsity of design points in the parameter space, and hence it provides an easier test for the calibration methods.

We defined the model output as $Y(\boldsymbol{\theta}, \mathbf{s}) = I(U(\boldsymbol{\theta}, \mathbf{s}))$ where

$$U(\boldsymbol{\theta}, \mathbf{s}) = \sqrt{1 - s_1^2 - \left(\frac{s_2}{1.5}\right)^2} - \theta_1 + \theta_2(s_2 + 1.5),$$

with $\mathbf{s} = (s_1, s_2)$ and $\boldsymbol{\theta} = (\theta_1, \theta_2)$ for $0.3 < \theta_1 < 0.65$, $0 < \theta_2 < 0.2$, $0 < s_1 < 1$, and $-1.5 < s_2 < 1.5$. We form an ensemble of model output with 100 runs where the design points for (θ_1, θ_2) are on a regular 10×10 lattice covering the entire parameter space $\Theta = [0.3, 0.65] \times [0, 0.2]$. Each model run $\mathbf{Y}(\boldsymbol{\theta}_i)$ is a spatial pattern on a 30×30 regular lattice that covers the entire spatial field $\mathcal{S} = [-1, 1] \times [-1.5, 1.5]$. By excluding the spatial locations \mathbf{s}_j having $Y(\boldsymbol{\theta}_i, \mathbf{s}_j) = 0$ for all $i = 1, \dots, p$, we reduce the number of spatial locations into $n = 608$. See Figure 6.2 for 20 sample model runs. The first parameter θ_1 controls the overall size of the grey area and the second parameter θ_2 changes both the shape and the size of the grey area. We set the number of principal components to be $J_y = 10$ based on leave-10-percent out cross-validation; using more than 10 principal components does not significantly improve the emulation performance in our cross-validation experiments. For a perfect model experiment we construct a synthetic observational dataset by choosing the model run with $(\theta_1, \theta_2) = (0.494, 0.089)$ as the synthetic truth and contaminating the corresponding model output with noise generated from the discrepancy process, defined as a Gaussian process with the partial sill of 1.5, the range of 0.03 and the nugget of 0.00001. This affects the resulting binary pattern considerably by changing roughly 10% of the original output. We repeat the same experiment with 20 different realizations from the discrepancy process and the results are summarized in Figure 6.3. The results show that on average we can recover the true input parameter setting reasonably well in this simple numerical setting.

6.5.2 Calibration of an Ice Sheet Model Using Binary Ice Patterns

We use a previously published model runs (Applegate et al., 2012) from the Greenland ice sheet model SICOPOLIS (Greve, 1997; Greve et al., 2011). Each model run is a high-dimensional spatial pattern on a 151×301 grid. Note that the model domain includes some areas covered by water where ice cannot accumulate, and hence there are grid points that contain no ice for all model runs. By excluding these locations, we obtain $n = 22,082$. The ensemble of runs is obtained by varying the following five different input parameters: ice positive degree-day (ice PDD),

snow positive degree day (snow PDD), geothermal heat flux, ice flow factor, and basal sliding factor. See Applegate et al. (2012) for a detailed description of the input parameters.

For the synthetic truth we choose model run #67 in Applegate et al. (2012), which was also used in Chang et al. (2013a) since (i) this run yields a modern ice sheet that is close to observations, and (ii) its parameter setting is not close to the edge of the cluster of design points and hence does not cause issues with extrapolation in emulation. Since one of the model runs is left out to construct the synthetic truth, the number of model runs used for emulation and calibration is $p = 99$. In principle we can calibrate the model using real world observations, but we choose not to provide real world projections here since we are not confident enough about the skill of the current Greenland ice sheet models including SICOPOLIS (Greve et al., 2011; Nowicki et al., 2013). Growing evidence indicates that the key assumption that these models are based on (“shallow ice approximation”) does not represent important characteristics of the Greenland ice sheet behavior (see, e.g. Little et al., 2007; O’Reilly et al., 2012).

The original model output consists of the spatial patterns of ice thickness. We dichotomize the spatial patterns at a certain thickness value to obtain binary spatial patterns. We use the cut-off of 100m, since in the SICOPOLIS ice with thickness of less than 100m are largely affected by the errors in the estimated modern temperature and precipitation fields and hence a cut-off value less than 100m may lead to unstable results.

To construct the synthetic observations we superimpose a discrepancy process to the original ice thickness pattern and dichotomize the result. This results in a binary observational dataset contaminated by spatially correlated discrepancy. In our experiment we subtract the same value of ice thickness (500m) from all model grid locations and dichotomize the resulting ice thickness pattern at 100m. This results in a smaller ice sheet than the ice sheet from the synthetic truth. We use this type of discrepancy to represent the situation where the ice sheet model is consistently overestimating the size of ice covered area, which is similar to the current state of the Greenland ice sheet models (Nowicki et al., 2013, their Figure 2). This overestimation by these models is mainly due to errors in the surface mass balance. We show the synthetic truth and the synthetic observational datasets

contaminated by the discrepancy in Figure 6.4.

We choose to use 10 principal components for dimension reduction based on the same cross-validation criteria described in the previous section. The leave-10-percent-out cross-validation confirms that our emulator can predict the model output precisely (Figure 6.5), with overall error rate of 6%. We give non-informative prior $IG(2, 3)$ to σ_d^2 and flat priors to the input parameters in $\boldsymbol{\theta}^*$. The range of each flat prior is restricted to the range for the design points $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p$. Using the standard Metropolis-Hastings algorithm we generate MCMC sample for $\boldsymbol{\theta}^*$ while integrating out σ_d^2 , $\boldsymbol{\eta}(\boldsymbol{\theta}^*, \mathbf{W})$ and v . For each experiment we obtain MCMC chain with 1,000,000 iterations. Looking at MCMC standard errors (Jones et al., 2006; Flegal et al., 2008) suggest that the sample size was large enough to estimate posterior means while accounting for autocorrelations. We also overlaid the marginal density plots obtained after half the run over the marginal density plots obtained after the entire run, confirming that the plots do not show significant changes and the results are therefore stable. Using the MCMC chain for the input parameters we can generate projections for volume change of the Greenland ice sheet from 2006 to 2100. We build an emulator for ice volume changes generated from the SICOPOLIS and convert the MCMC chain for input parameters into projections using the emulator. The computation takes about 10 hours for the entire MCMC run on a system with Intel Xeon E5450 Quad-Core 3.0 GHz..

The pairwise joint densities of input parameters are shown in Figure 6.6. The density plots indicates that the marginal posterior density for the ice PDD factor peaks around the true parameter values, but the joint densities for other parameters do not recover the true values. We hypothesize that the binary spatial patterns are not informative about parameters other than the ice PDD. This result is consistent with earlier observations that surface mass balance controls the distribution of glacier ice (e.g. Greve, 1997). Moreover, the marginal densities for the snow PDD factor and the basal sliding factor show bimodality and have peaks at the points that are not close to the true parameter values. Understanding the cause of this phenomenon is not straightforward and therefore requires further careful investigation.

The result in Figure 6.7 shows the predictive distribution of the ice volume change projections corresponding to the estimated posterior density. For compari-

son we have also computed two additional predictive densities by (i) assuming that all the parameter settings in the model ensemble are equally likely and (ii) using a filter that chooses the parameter settings which result in ice covered area within 10% of the observed area. The results show that our calibration approach results in smaller prediction uncertainty compared to the other two simple approaches.

6.6 Discussion

In this chapter we have proposed a computer model emulation and calibration approach for binary spatial data, which are useful for ice sheet model calibration problems. To our knowledge, this is the first attempt to formulate a computer model calibration approach using non-Gaussian spatial data. Using the logistic principal component analysis and basis representation, our approach can efficiently handle high-dimensional non-Gaussian spatial data. The logistic principal component analysis is a special case of principal component analysis for the one parameter exponential family, and therefore it can be easily extended to other types of spatial data such as spatial count processes. The approach can be even further extended to the case where the variables in model output and observational data have different distributions in the one-parameter exponential family.

The simple numerical experiment shows that our approach can estimate the input parameters reasonably well even with the presence of considerable model-observation discrepancy. For the ice sheet model calibration problem, our results show identifiability issues in inferring the input parameters except for the ice PDD, but it can still make accurate projections for the ice volume changes. We hypothesize that this is because (i) the effect of the other four input parameters on the shape of the ice sheet is not strong enough to be detected under the presence of severe discrepancy, and (ii) the future projection of ice volume change is not significantly affected by those parameters over the next hundred years.

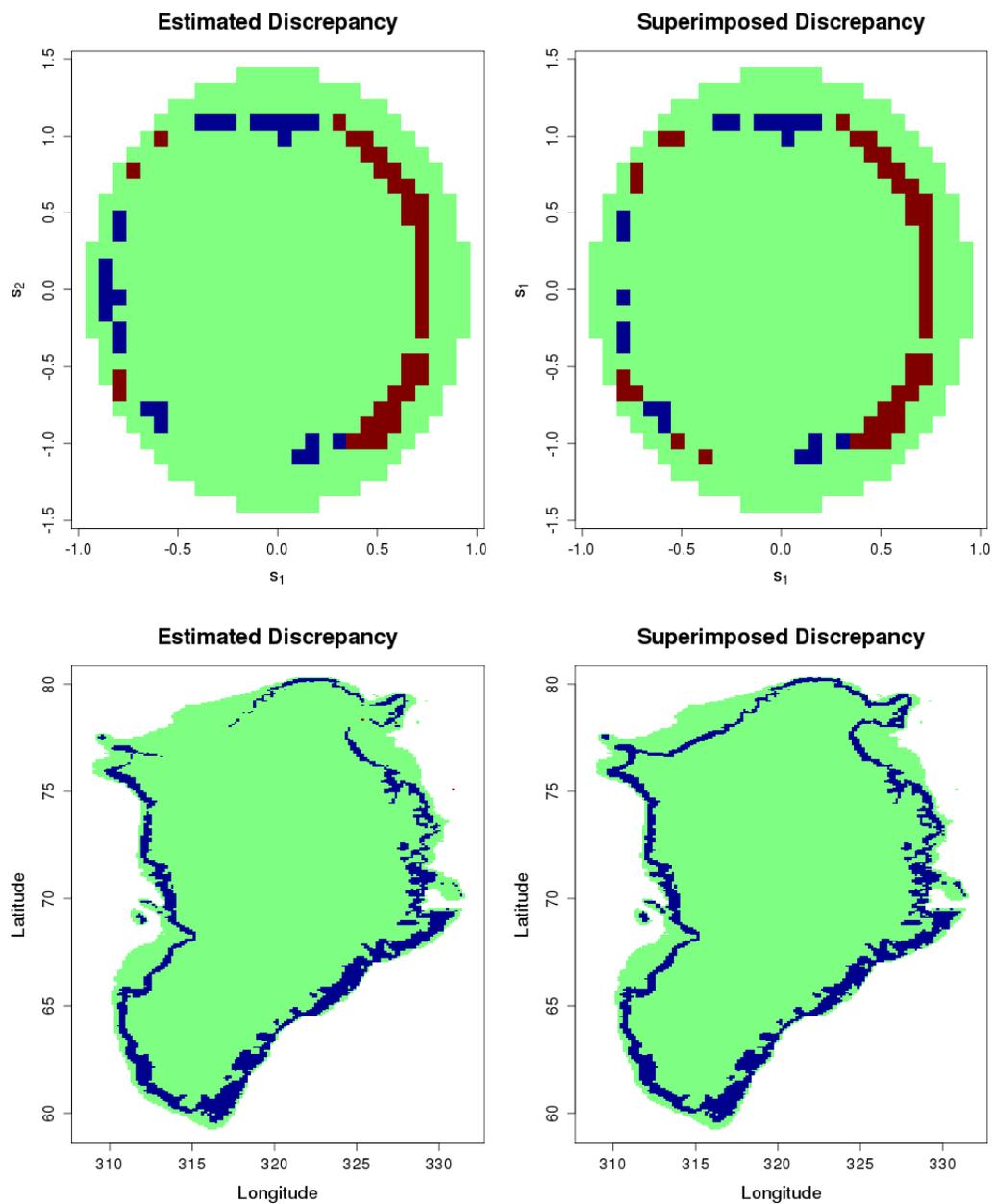


Figure 6.1: Comparison of the estimated discrepancy patterns by the procedure described in 6.4.2 (left panels) and the true discrepancy patterns that are actually used in 6.5 (right panels). The upper panels show the results for one of the discrepancy patterns used in the simple example in 6.5.1. The other 19 cases with different realizations of the discrepancy process exhibit the similar results. The lower panels show the discrepancy patterns in the SICOPOLIS calibration problem in Section 6.5.2. In both cases the estimated discrepancies well capture the overall patterns of the true discrepancy patterns.

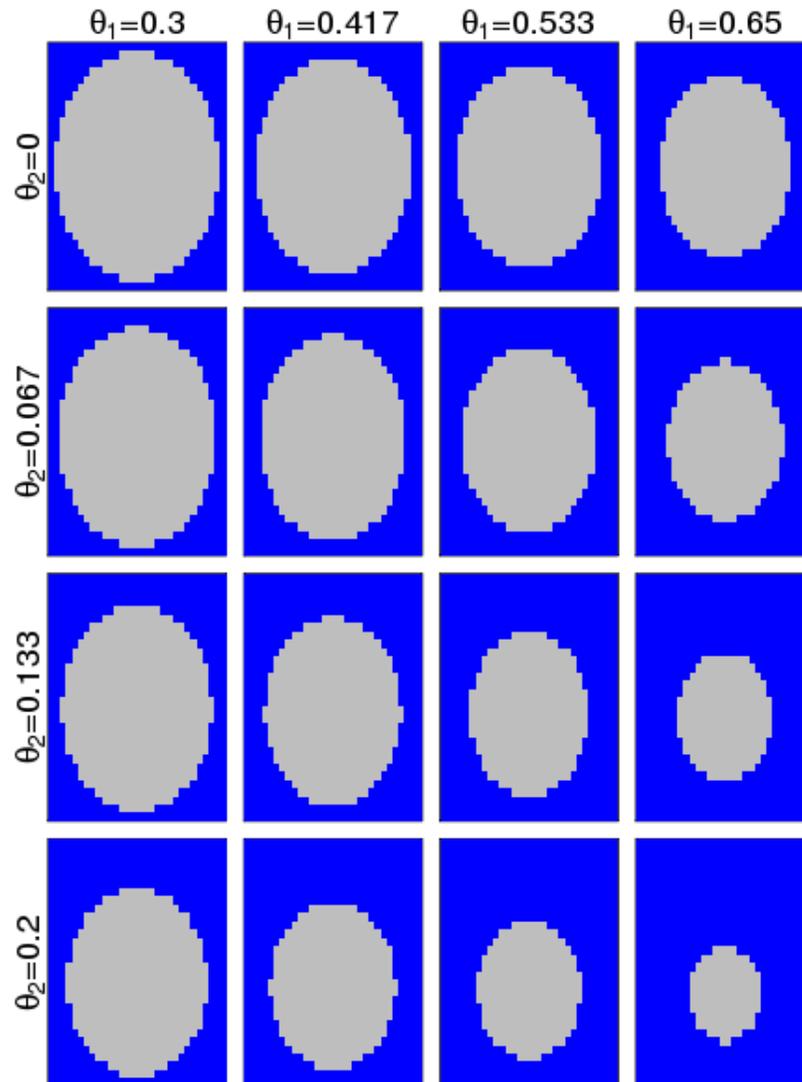


Figure 6.2: Example of 20 model runs out of total 100 model runs. The resolution of each image is 30×30 . The first parameter θ_1 controls the overall size of the grey pattern and the second parameter θ_2 changes both the shape and the size of the pattern.

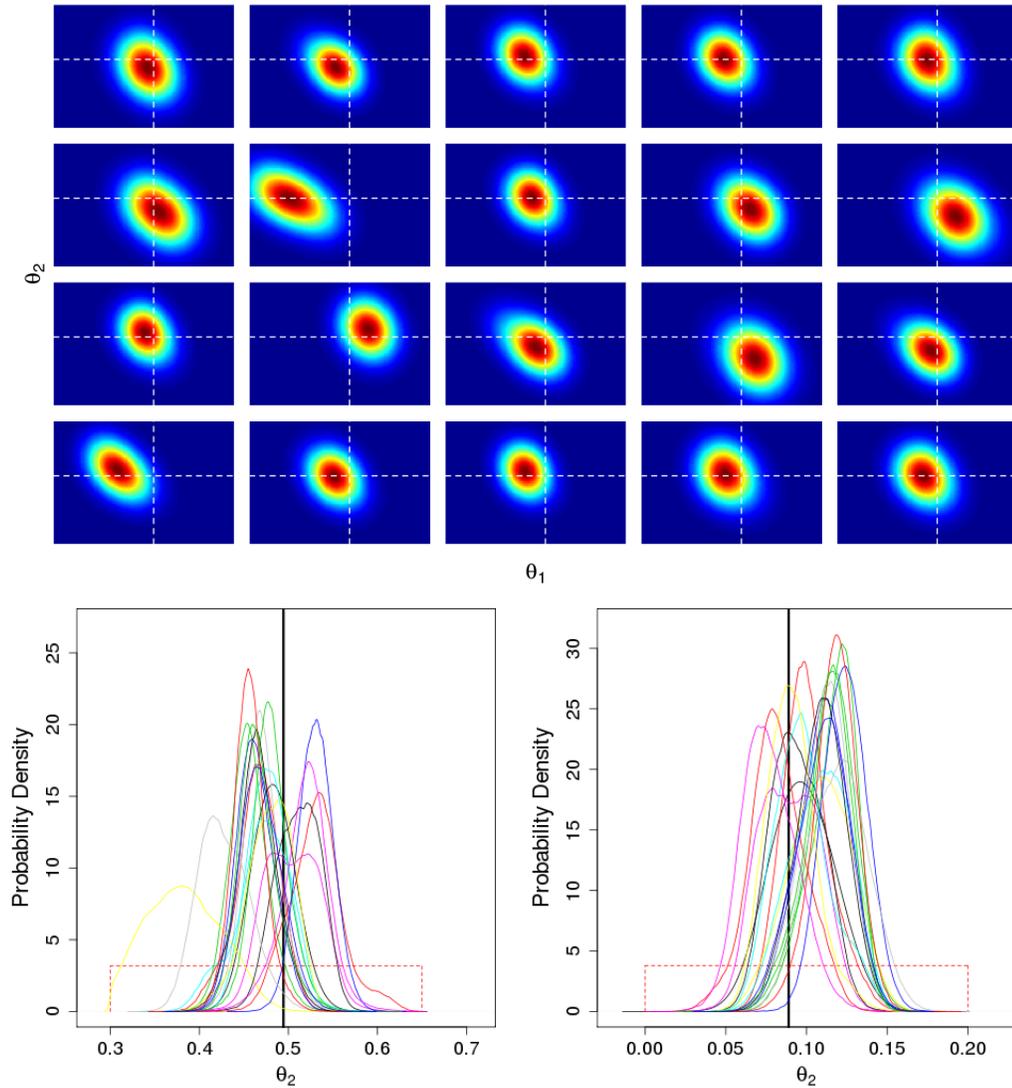


Figure 6.3: The joint (upper panel) and the marginal (lower two panels) posterior densities of the input parameters based on synthetic observational data sets with 20 different realizations from the discrepancy process defined in 6.5.1. Most of the cases the resulting posterior densities recover the true input parameter setting (white dashed lines in the upper panel and black solid line in the lower panels) reasonably well. The marginal posterior distributions have significantly smaller variation comparing to the prior distribution, showing that the parametric uncertainty has been significantly reduced by calibration.

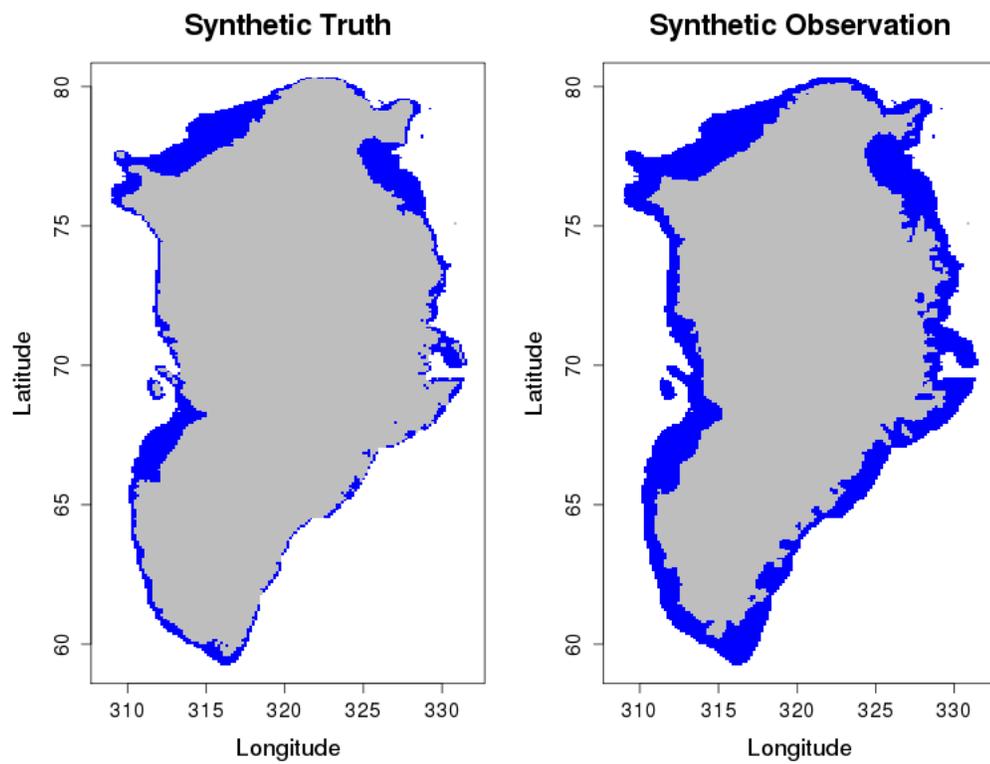


Figure 6.4: The output from the original model run chosen as the synthetic truth (left panel) and the synthetic observational data obtained by superimposing the discrepancy term described in Section 6.5.2. The synthetic observation has a notably smaller ice covered area.

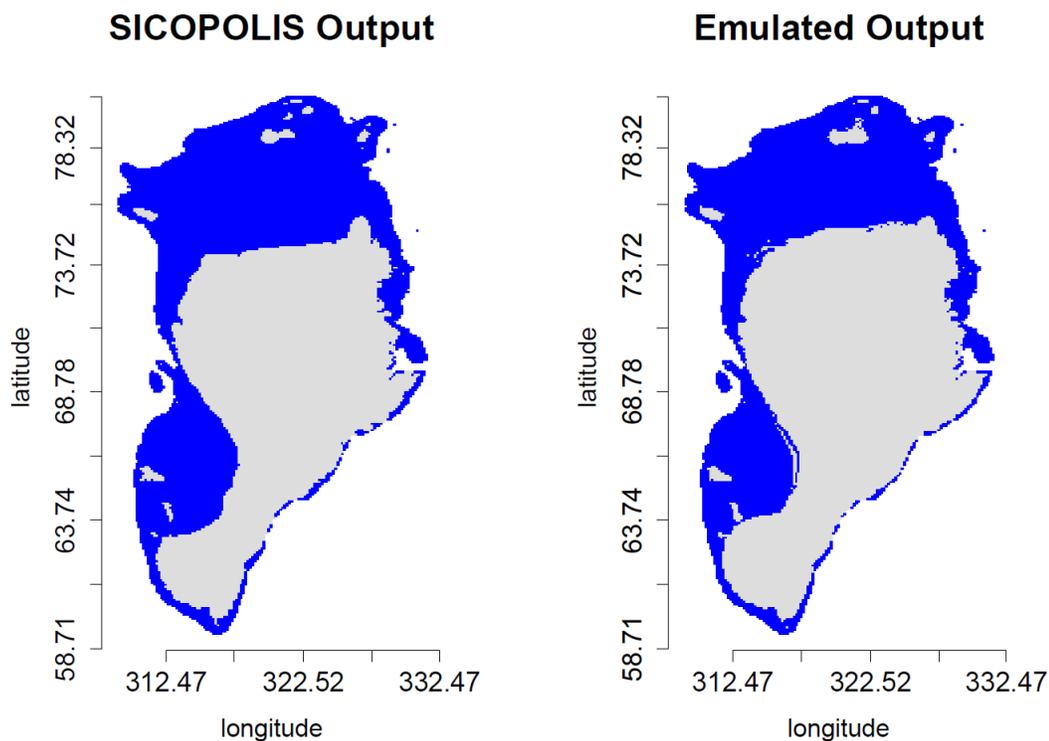


Figure 6.5: An example of leave-10-percent-out cross-validation result to verify the prediction performance of our emulator based on logistic PCA. The result shown here is for model run #74 in Applegate et al. (2012). The graphical comparison shows that the emulated model output (right panel) predicts the original SICOPOLIS model output precisely. Similar results hold for the other model runs.

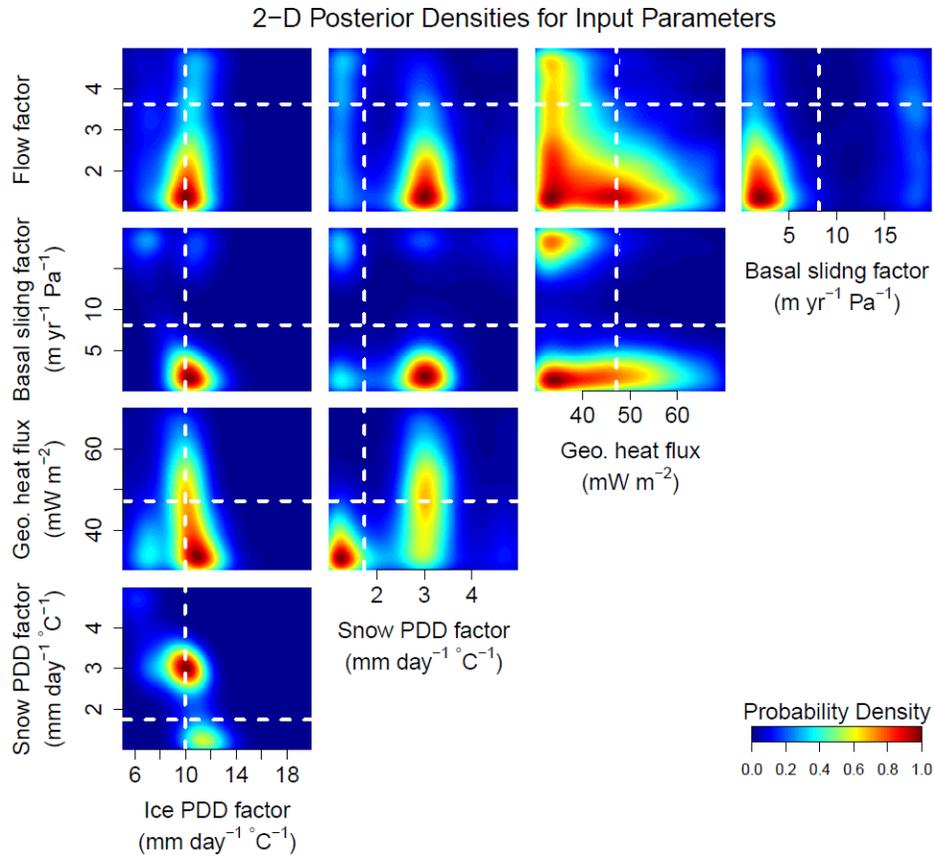


Figure 6.6: The pairwise joint posterior densities for the input parameters from the calibration results in Section 6.5.2. The white dashed lines show the input parameter settings for the synthetic truth. Only the marginal posterior density of the ice PDD factor peaks around the values for the synthetic truth. The marginal densities for the snow PDD factor and the basal sliding factor show bimodality, which requires further investigation.

Illustrative projections based on synthetic data

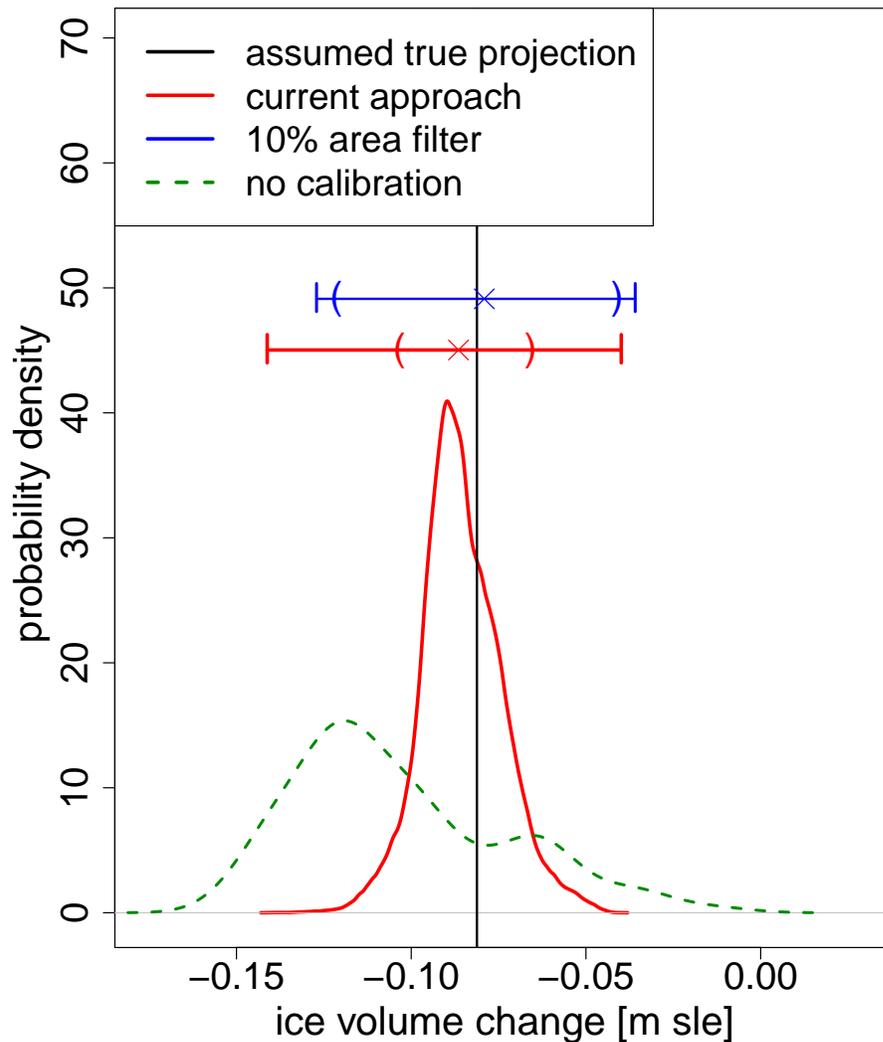


Figure 6.7: The ice volume change projections from 2006 to 2011 generated by the SICOPOLIS Applegate et al. (2012) based on our calibration approach (solid red line), the 10% area filter that chooses model runs within 10% of the observed ice covered area (solid blue line), and no calibration (dashed green line). The parentheses on the horizontal bars represent 95% prediction interval and the vertical bars the ranges. Our approach yields a narrower 95% prediction interval than the filtering approach. The predictive distribution from our approach shows slight bias probably due to the biases in the posterior densities for the input parameters other than the ice PDD (see Figure 6.6).

Discussion

7.1 Summary and Contributions

In this thesis I have developed computationally efficient computer model calibration approaches and applied them to climate and ice model calibration problems. I have formulated computer model calibration frameworks for high-dimensional spatial data based on two different approaches: dimension reduced calibration and block composite likelihood. The results from the climate model calibration examples show that these approaches can handle the large spatial data with high computational efficiency and infer the input parameters in a statistically sound manner. I have extended the reduced-dimension approach to computer model calibration using binary spatial datasets and applied the method to an ice model calibration problem. The perfect model experiment results indicate that the calibration using binary spatial patterns results in a better projection for future ice status and the corresponding sea level rise projection.

In Chapter 4 I have introduced a highly computationally efficient calibration approach based on principal component analysis and kernel convolution. Using this method I have calibrated a climate model based on spatial data with tens of thousands of spatial locations. The method scales, in principle, to much larger data sets because the principal components approach reduces potentially massive spatial data sets to low dimensional components that can be emulated independently. I have compared the calibration results based on spatial data with different level of aggregation and discovered that calibration using unaggregated data results in the

smallest parametric uncertainty. The result has important implications for climate uncertainty quantification; the common practice of data aggregation in climate science research may cause information loss in parameter calibration and results in a larger parametric and projection uncertainty. Therefore, using the proposed calibration approach that can efficiently handle the high-dimensional spatial data can lead to a better climate uncertainty quantification.

In Chapter 5 I have proposed a calibration method based on block composite likelihood. This is, to my knowledge, the first application of composite likelihood to a computer model calibration problem. While relatively less efficient in computation comparing to the reduced-dimension approach, this method enables rigorous approximation for the likelihood function and asymptotic analysis on the posterior estimates. This study also opens up the possibility for application of other types of composite likelihood and spatial blocking-based approaches to calibration problem.

In Chapter 6 I have generalized the reduced-dimension calibration framework (presented in Chapter 3) to the one-parameter exponential family and formulated a computationally efficient calibration framework using high-dimensional non-Gaussian spatial data. This approach is particularly useful for problem of calibrating ice sheet models, where computer model output and observational data are non-Gaussian spatial fields. There are some previous results showing that calibration using spatially resolved information can significantly reduce the parametric uncertainty in ice volume change projections (Chang et al., 2013a), but ice sheet model calibration is still largely unexplored area that still has significant room for improvement. The perfect model experiment results show that the proposed approach can resolve the computational and inferential challenges for calibrating a Greenland ice sheet model using high-dimensional binary spatial fields from model runs and observations.

7.2 Caveats and Potential Improvements

Although the proposed methods here provide computationally efficient approaches to climate model calibration using high-dimensional spatial datasets, they share several common caveats of previously studied computer model calibration and

spatial modeling approaches. One common issue across the proposed calibration approaches is the assumption of separable covariance structure. In the calibration examples in Chapter 4 and 6, the climate models have multiple input parameters and the covariance structure defined for the input parameter space is based on the separability assumption between the input parameters. When the input parameters have strong interactions with each other, the separability assumption may result in a poor approximation to the covariance structure for the input parameters. To my knowledge there has not been any work that attempts to address this issue. Similarly for the composite likelihood approach we need to assume separability between space and input parameters for computational tractability and difficulty in defining a sensible non-separable covariance structure. In the simulated examples in Chapter 4-6 the separability assumption seems to provide a good approximation to the covariance structure for the climate model outputs, but there might be cases where the separability assumption results in a poor emulation for the climate model. One may consider applying non-separable covariance structures developed for modeling space-time processes (see e.g. Cressie and Huang, 1999; Gneiting, 2002; Stein, 2005), but adapting those approaches for computer model calibration is not straightforward.

Another possible issue is information loss caused by dimension reduction used in the approaches in Chapter 4 and 6. It is well known that reduced dimensional approaches cause loss of information for smoothness of Gaussian processes (Stein, 2014). This is usually not a problem when the main purpose is to capture the large scale trend and/or to model a smooth process with a large nugget effect, which were the case for my applications in those chapters. However, there might be a case where the computer model of interest may yield non-smooth spatial processes as their output and therefore reduced dimension approaches are not suitable. In such case, the composite likelihood calibration framework described in Chapter 5 may provide a better approach.

Finally, the calibration framework in Chapter 6 is not applicable to non-Gaussian distributions that do not belong to the one-parametric exponential family. Therefore, our approach does not apply for calibration problems based on important climate processes such as the spatial pattern of precipitation or ice thickness, both of which are in the form of zero-inflated continuous data. Modelling this

type of processes requires a latent Gaussian process approach that poses significant computational challenges in integrating out the latent process (De Oliveira, 2005). The stochastic expectation-maximization algorithm (see, e.g. Diebolt and Ip, 1996; Nielsen, 2000) or approximate Bayesian computation (see, e.g. Tavaré et al., 1997; Beaumont et al., 2002) might provide a solution for this problem, but for such algorithms we need to be able to efficiently generate the latent processes in a repeated manner.

There are also several caveats in my scientific results. First, for the calibration examples in Chapters 4 and 5 the calibration results are based on only one observational data set and the calibration results may have considerable biases due to the unresolved natural variability (Olson et al., 2013). This is a common issue for most calibration results based on only one source of observational data, and using multiple observational datasets from different sources may resolve this issue (Bhat et al., 2012). Second, the model output and the observational data are averaged over a certain time period and the time component is not included in the emulation and calibration model. It is not clear whether including the time component in the calibration model explicitly improve the calibration result due to the increased information or introduce a bias due to the effect of unresolved natural variability. Third, most of the modern ice models including the ice sheet model used in Chapter 6 have several limitations (Greve et al., 2011; Nowicki et al., 2013), and therefore we choose not to provide projections for sea level rise based on observational data. This leaves the problem of making realistic projections for future ice volume change using more advanced models as a future research.

7.3 Avenues for Future Research

In addition to addressing the specific issues listed above there are several directions for future research. Here I focus on four potential ideas among these possibilities: a hierarchical calibration model using spatial blocks, sufficient dimension reduction for input parameter space, a calibration approach with multiple types of distributions, and an emulator that operates simultaneously in parameter space and in scenario space.

- The block composite likelihood approach described in Chapter 5 uses spa-

tial blocking only to reduce the computational cost for likelihood evaluation. It is possible to expand the idea of spatial blocking to introduce a more flexible spatial model in climate model calibration. In a similar manner to the composite likelihood formulation in Caragea and Smith (2006) one can formulate a hierarchical model where its first layer models the dependence between the blocks and the second layer models the dependence within each block. Moreover, by using different parameter values for different blocks one can formulate a calibration framework that allows non-stationary spatial processes. Unlike the approach described in Chapter 5 where the shapes of blocks have virtually no effect on the calibration results due to the stationarity, results from the hierarchical model are largely affected by the shapes of the spatial blocks. Therefore, in this framework it is important to find an optimal blocking that results in the maximum parametric uncertainty reduction. One possible approach to this problem is the treed Gaussian process model (Gourieroux et al., 2008).

- Input parameters for climate and ice models often show strong interactions with each other (see, e.g. Chang et al., 2013a) and therefore the model output depends on the input parameters only through some low dimensional functions of them. In such case sufficient dimension reduction (see, e.g. Cook, 1998; Adraghi and Cook, 2009; Li, 1991) for input parameter space may reveal those low dimensional functions and inform us on how the parameters are interacting and which parameters are influential on the outcome (Cook, 2004). Such information is especially useful for selecting parameters for further investigation and designing sampling schemes for follow-up studies.
- In Chapter 6 we introduce a calibration approach using non-Gaussian spatial data. Since the dimension reduction is applied to the natural parameters, it is possible to extend the method to a calibration approach that can handle multiple distributions from the one-parameter exponential family. Such method is especially useful for climate model or ice model calibration by incorporating various sources of information from modern observations such as remote sensing data and paleoclimate data such as ice core observations.
- The emulator that we have developed in this thesis interpolates the model

runs only in the parameter space. Building an emulator that interpolates model runs with different forcing scenarios is of great interest to the climate research community; there has been some recent work in this regard (Castruccio et al., 2013b,a). It may be possible to extend my approaches to build an emulator that can simultaneously interpolate the climate model output across different input parameter settings and forcing scenarios. In addition to the computational challenges for dealing with large space-time data, this problem poses additional modeling challenges to define a sensible covariance structure for model runs with different input parameters and different forcing scenarios.

- Finally, it is worth noting again that dimension reduction methods as well as composite likelihood methods are very widely applicable to problems that are neither spatial nor related to computer model emulation and calibration. Both methodologies represent general approaches for tackling the immense computational challenges posed by large data sets and increasingly sophisticated models; developing new methods and theory based on these approaches in both Bayesian and non-Bayesian settings is an interesting and potentially very important avenue for future research.

Reduced-Dimensional Calibration.

A.1 Computational Cost and Storage Issue in Emulation Problem

For the 3-D case, the computational cost for matrix computation in a single likelihood evaluation is approximately $\frac{1}{3} \times 250^3 \times 61051^3 = 1.185 \times 10^{21}$ flops, which takes more than ten million hours for a high performance modern single-core processor, typically having 3.0 GHz. Note that this computational time assumes sufficient memory is available; when it is not, as will often be the case, computation could be infeasible or slowed down even further due to memory swapping issues. In our application, a covariance matrix with $p^2 \times n^2 = 250^2 \times 61051^2$ elements needs to be stored for computation, which requires $8 \times \frac{250^2 \times 61051^2}{1024^3} = 1,735,624$ Gb of storage space on double precision. This costs about 1.7 million US dollars at the current price of 10 dollars per gigabyte.

A.2 Covariance Structure Implied by Separability Assumption

For the i th principal component process $w_i(\cdot)$, the covariance between two parameter settings $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ is given by

$$\begin{aligned}
 \text{Cov}(w_i(\boldsymbol{\theta}_1), w_i(\boldsymbol{\theta}_2)) &= \text{Cov}\left(\int Y(\boldsymbol{\theta}_1, \mathbf{s}_1)\mathbf{e}_i(\mathbf{s}_1)d\mathbf{s}_1, \int Y(\boldsymbol{\theta}_2, \mathbf{s}_2)\mathbf{e}_i(\mathbf{s}_2)d\mathbf{s}_2\right) \\
 &= \int \mathbf{e}_i(\mathbf{s}_1)\left(\int \text{Cov}(Y(\boldsymbol{\theta}_1, \mathbf{s}_1), Y(\boldsymbol{\theta}_2, \mathbf{s}_2))\mathbf{e}_i(\mathbf{s}_2)d\mathbf{s}_2\right)d\mathbf{s}_1 \\
 &= C_\theta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\int \mathbf{e}_i(\mathbf{s}_1)\int \text{Cov}(\mathbf{s}_1, \mathbf{s}_2)\mathbf{e}_i(\mathbf{s}_2)d\mathbf{s}_2d\mathbf{s}_1 \\
 &= C_\theta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\lambda_i.
 \end{aligned}$$

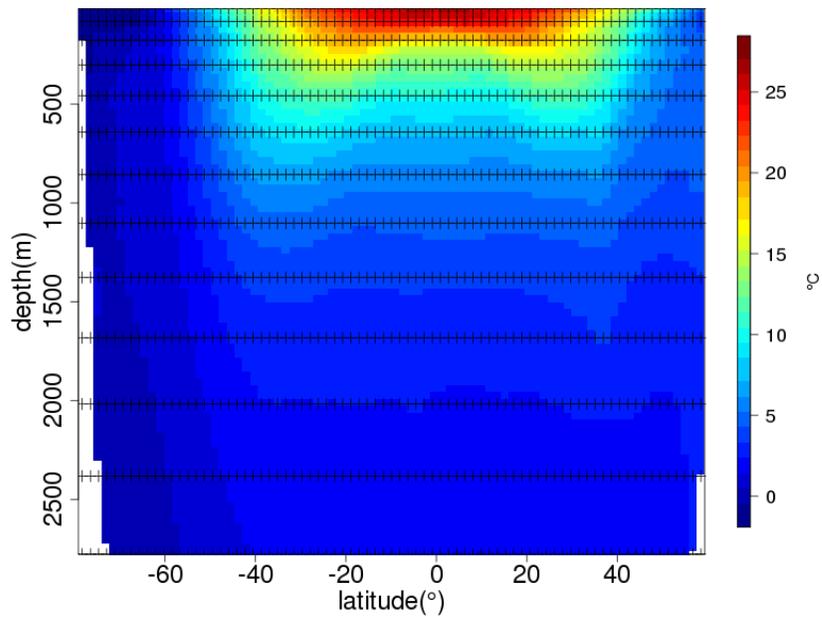
where λ_i is the corresponding eigenvalue.

A.3 Diagnostics for Prediction Performance of PC Emulator

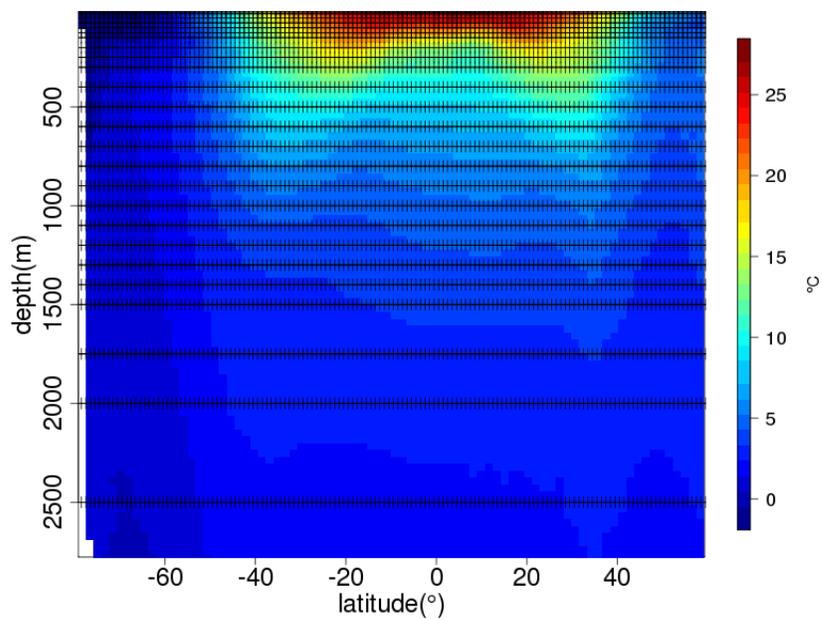
To test the performance of our emulator, we computed uncorrelated standardized prediction errors (Bastos and O'Hagan, 2009) using pivotal Cholesky decomposition for the first four principal components at five design points to have 20 predictions in total. If the emulators predict the processes of principal components reasonably well, the errors should be approximately independent and follow a t -distribution with 5 degrees of freedom. The diagnostic plots based on the computed errors are shown in Figure A.5. The graphical diagnostics indicate that our emulators for principal components predict the principal components with reasonable precision. Gaussian process emulation does not perform as well as we go beyond the first 4 leading principal components; this is not surprising because these components tend to be noisy. Note, however, that this does not cause any problems because the contribution from the non-leading principal components to the final result is minimal.

A.4 Computing Time Reduction by Parallel Computing

It is possible to further reduce the computing time for our approach using parallel computing, which may be conveniently carried out in R for instance by using the **snow** package (Tierney et al., 2013). For the 3-D case, we found that distributing the computation load on 4 processors reduces the overall computing time by 40%. The reduction seems less significant due to the fact that there is significant overhead due to communication time among the processors, taking up about 35% of the overall computing time. We note, however, that this is largely because the computation is already very fast without parallelization (less than 3 minutes). For problems requiring much longer computing time due to large p and/or J_y , parallelization will provide a more significant reduction in computing time.

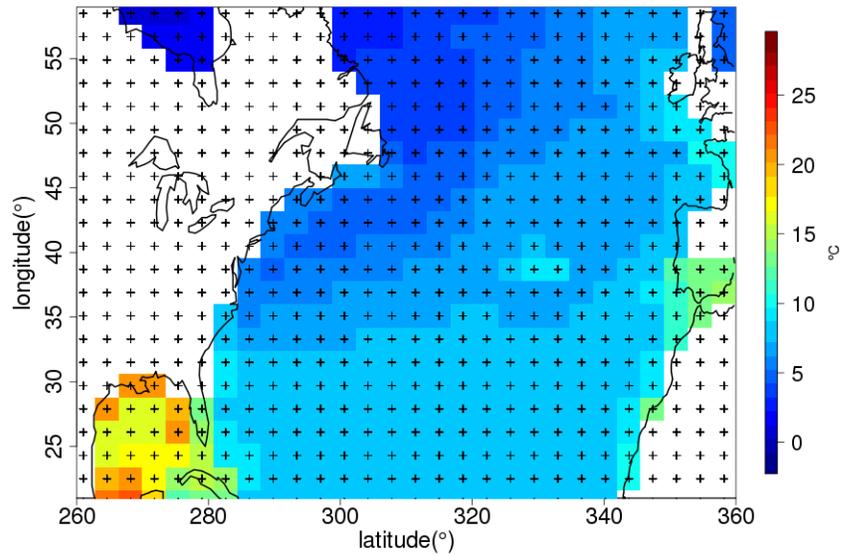


(a) UVic ESCM grid

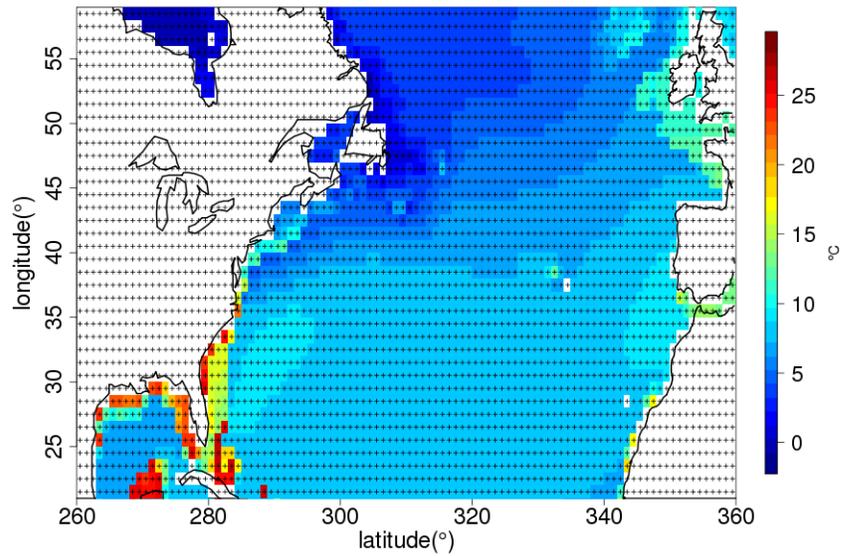


(b) World Ocean Atlas grid

Figure A.1: Comparison of UVic ESCM grid (Weaver et al., 2001) and World Ocean Atlas grid (Antonov et al., 2010; Levitus et al., 2012), projected on latitude-depth space. The pattern embedded in each plot is the mean ocean temperature averaged over longitude (0-360°) and time (1955-2006). Note that the temperature pattern shown in (a) is an example of 250 model runs described in Section 2.



(a) UVic ESCM grid



(b) World Ocean Atlas grid

Figure A.2: Comparison of UVic ESCM grid and World Ocean Atlas grid, projected on longitude-latitude space. For better visibility of the figures we display only the North Atlantic area, but the model output and observations cover the entire globe. The pattern embedded in each plot is the mean ocean temperature averaged over depth (0-3000m) and time (1955-2006). Note that the temperature pattern shown in (a) is an example of 250 model runs described in Section 4.2.

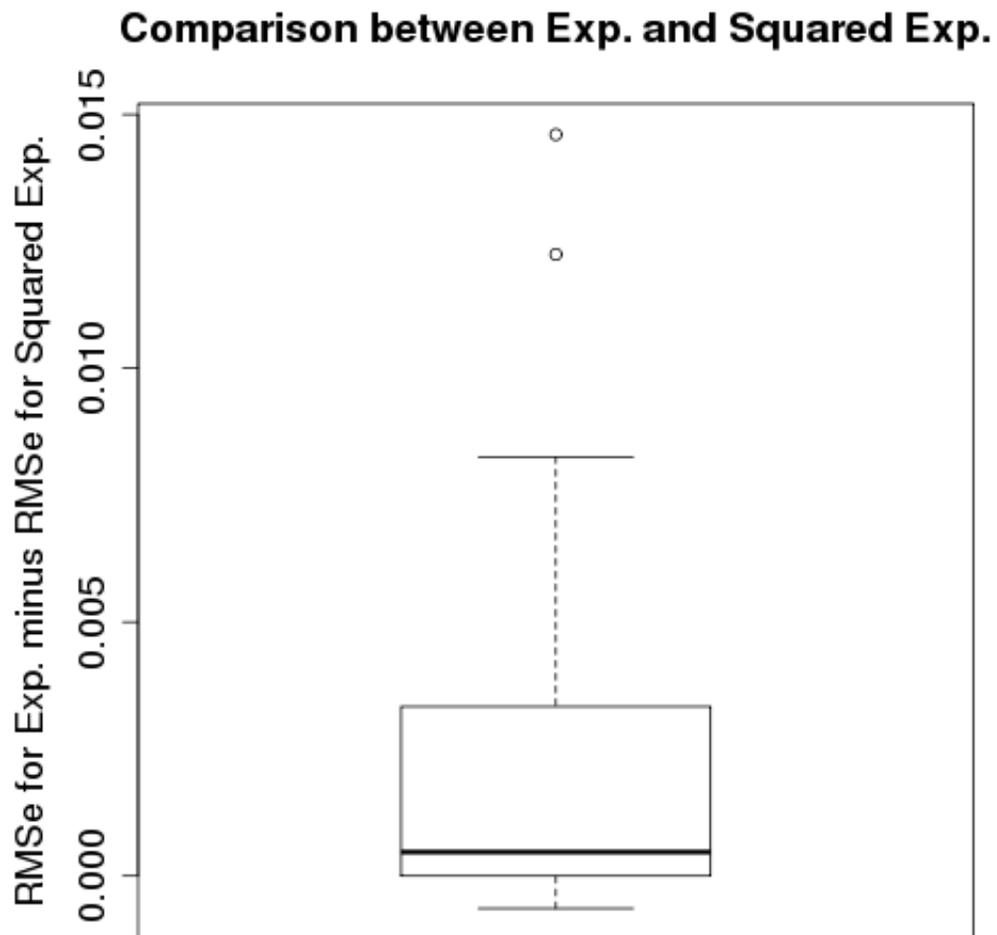


Figure A.3: Comparison of emulation performance between the exponential covariance function and the squared exponential covariance function based on 50 different leave-10-percent-out cross validation scenarios. In each scenario, we randomly hold out 25 different 3-D model outputs and predict them using the remaining 225 model runs. The root mean squared errors (RMSe) are computed for each method, and the RMSe for the exponential covariance function is subtracted from the RMSe for the squared exponential function. The boxplot of differences in RMSe shows that squared exponential covariance function outperforms or shows similar performance to the exponential covariance function. Any positive RMSe corresponds to squared exponential covariance outperforming exponential covariance.

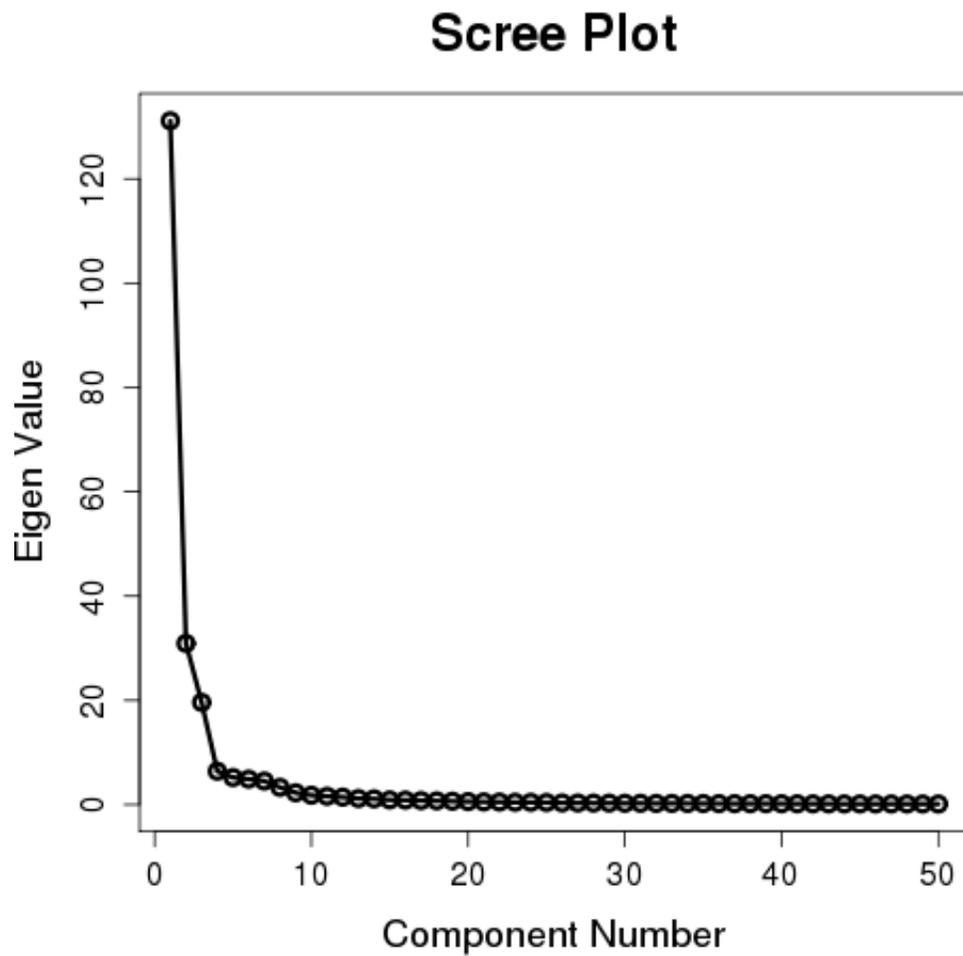
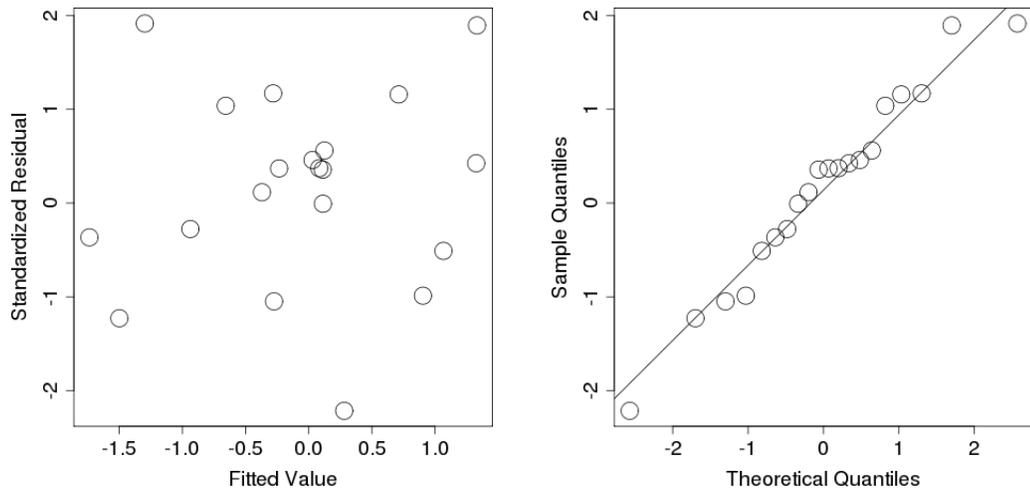


Figure A.4: Scree plot showing the eigenvalues for the 3-D case. The magnitude of the eigenvalues decreases rapidly for the first 10 principal components, and hence the remaining eigenvalues are extremely small comparing to the leading eigenvalues.



(a) Residual vs. fitted values

(b) t Q-Q plot, with degrees of freedom 5

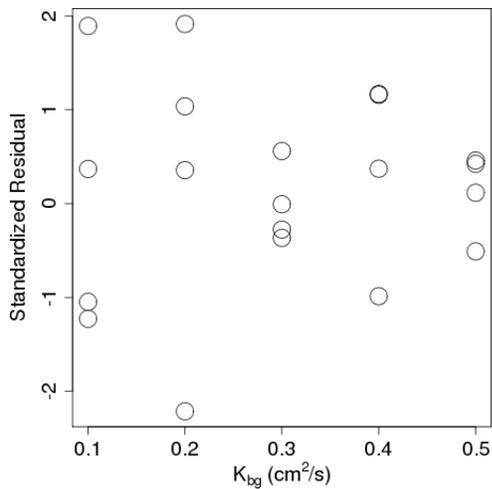
(c) Residual vs. input parameter (K_{bg})

Figure A.5: Graphical diagnostics using standardized uncorrelated prediction errors. For the first four principal components, prediction errors using pivotal Cholesky decomposition are computed at five K_{bg} settings (0.1, 0.2, 0.3, 0.4, 0.5) while the other two parameters are fixed at $A_{sc} = 1.5$ and $C_s = 3.976$, resulting in 20 uncorrelated standardized prediction errors.

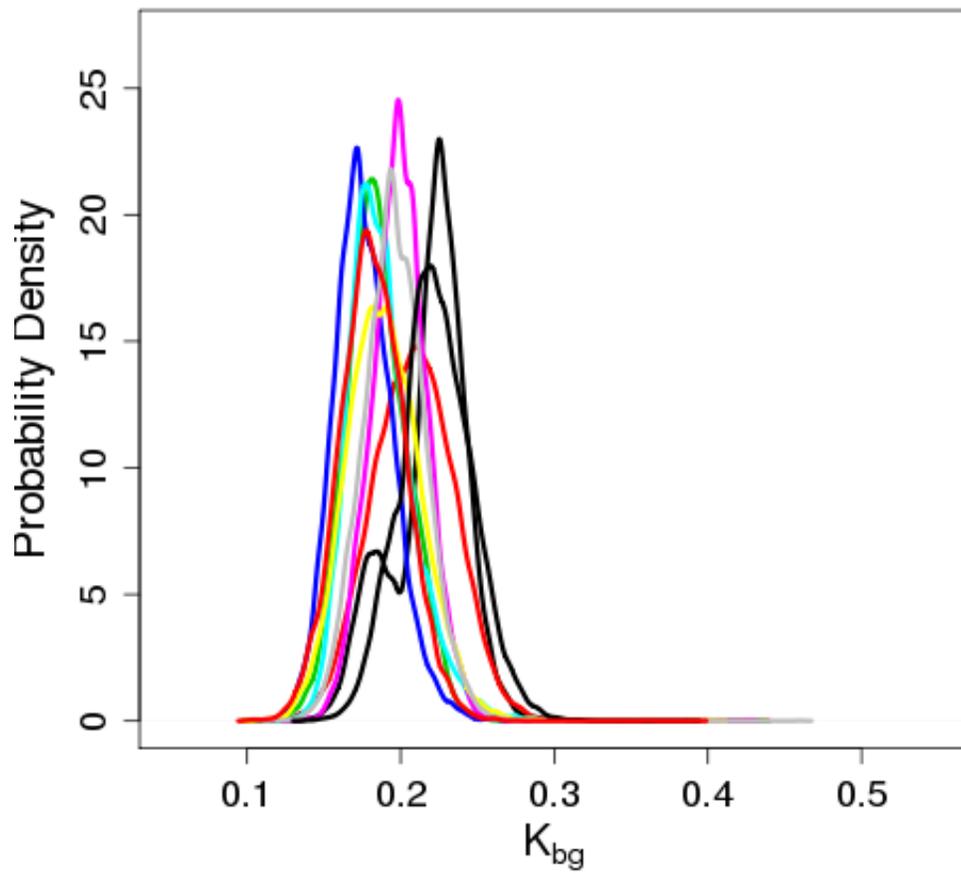


Figure A.6: Posterior densities of K_{bg} based on 10 different random samples from 3-D locations. Each curve represents the posterior density from one sample with size $n = 1,300$.

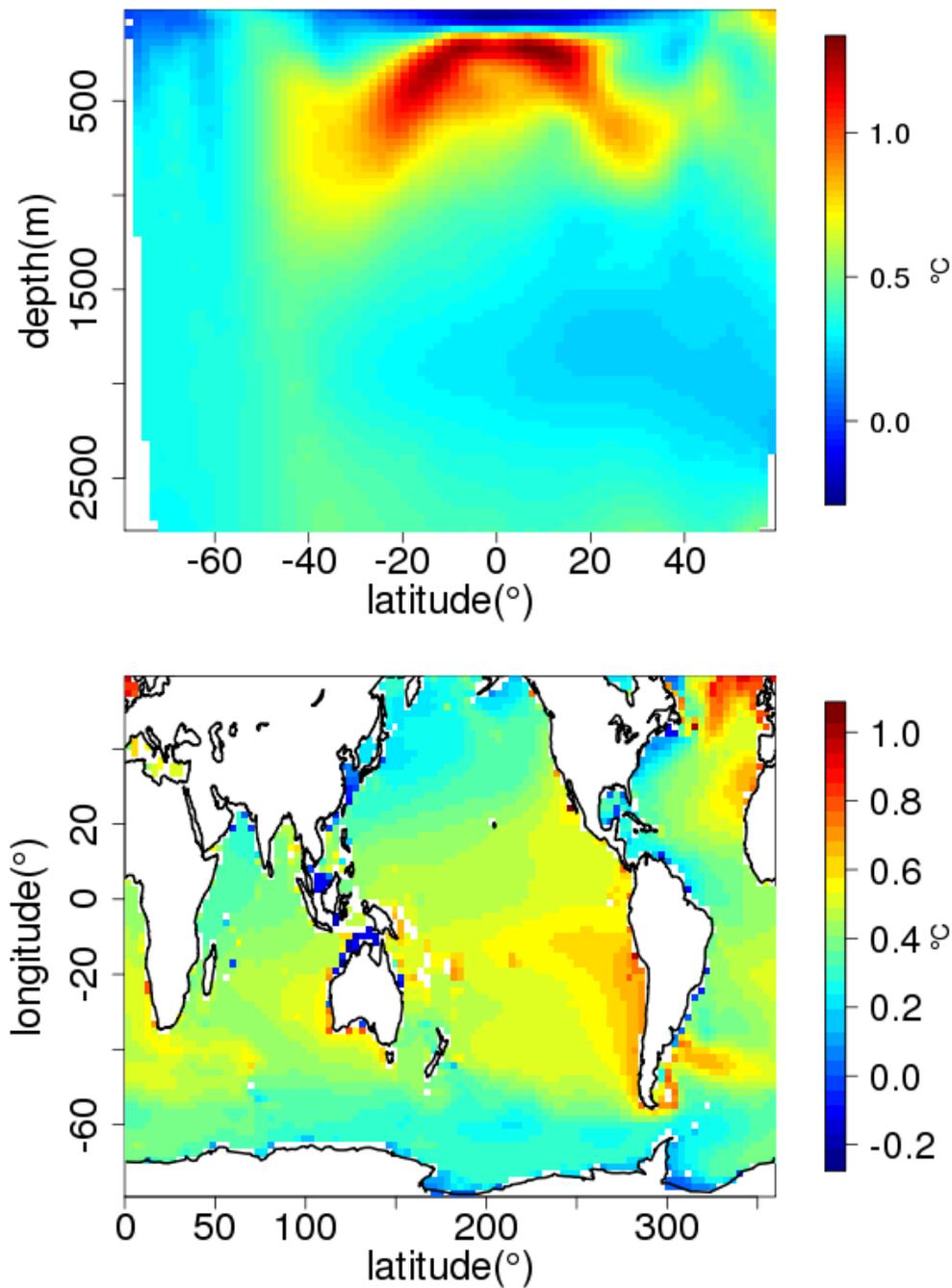


Figure A.7: Spatial patterns for the first principal component basis vector. The latitude-depth pattern seems to show the contrast between the near the surface part and the deeper part of the ocean in the low latitude area caused by the effect of K_{bg} on diffusion (Kriegler, 2005). The latitude-longitude pattern seems to show the effect of K_{bg} on AMOC in the high-latitude North Atlantic, Western Boundary Currents (primarily Kuroshio current), and Eastern Boundary Currents (e.g., Peru current).

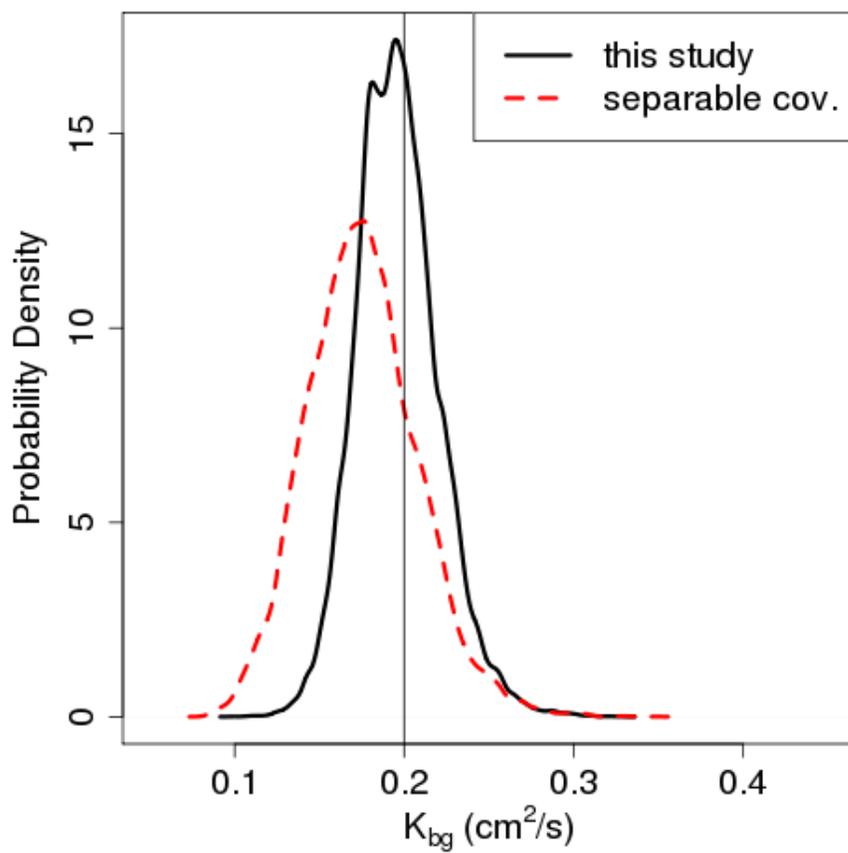


Figure A.8: Comparison of posterior densities of K_{bg} computed by a) our PCA-based approach (black solid line) and b) assuming separable covariance structure (red dashed line). The results are based on 2D (latitude-depth) ocean temperature patterns.

Composite Likelihood Calibration

B.1 Computation of \mathbf{P}_n and \mathbf{Q}_n

In this supplementary material, we describe the matrix computation for $\mathbf{P}_n = \text{Cov}(\dot{cl}_n(\boldsymbol{\psi}))$ and $\mathbf{Q}_n = \text{E}(\ddot{cl}_n(\boldsymbol{\psi}))$. For ease of computation, it is useful to rewrite the composite likelihood function when $p = \infty$ as

$$cl_n(\boldsymbol{\psi}) \propto -\frac{1}{2} \left(\log |\Sigma^{\bar{\mathbf{Z}}}| + (\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*)^T (\Sigma^{\bar{\mathbf{Z}}})^{-1} (\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*) \right) - \frac{1}{2} \left(\sum_{i=1}^M \log |\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}| + \sum_{i=1}^M (\mathbf{Z}_{[i]} - \mathbf{Y}_{[i]}^*)^T \mathbf{A}_i^T (\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}})^{-1} \mathbf{A}_i (\mathbf{Z}_{[i]} - \mathbf{Y}_{[i]}^*) \right),$$

where \mathbf{A}_i is a $(n_i - 1) \times n_i$ matrix such that

$$\mathbf{A}_i = (\mathbf{I}_{(n_i-1) \times (n_i-1)} \mathbf{0}_{(n_i-1) \times 1}) - \mathbf{a}_i \left(\frac{1}{n_i}, \dots, \frac{1}{n_i} \right)_{1 \times n_i},$$

and \mathbf{a}_i is a $(n_i - 1) \times 1$ vector such that

$$\mathbf{a}_i = (\zeta_{\theta} \gamma^{(i)} + \lambda^{(i)}) \left\{ \Sigma^{\bar{\mathbf{Z}}} \right\}_{ii}^{-1}.$$

$\mathbf{Z}_{[i]}$ is a $n_i \times 1$ vector containing all the n_i observational data in the i th spatial block without omission, and $\mathbf{Y}_{[i]}$ is a $n_i \times 1$ vector of model output at $\boldsymbol{\theta}^*$ defined in the same way. Omitting the part irrelevant to the data, the partial derivative

of $cl_n(\boldsymbol{\psi})$ with respect to the j th computer model parameter, θ_j^* , is given by

$$\frac{\partial cl_n(\boldsymbol{\psi})}{\partial \theta_j^*} \propto \bar{\mathbf{B}}_j^* (\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*) + \sum_{i=1}^M \mathbf{B}_{i,j}^* (\mathbf{Z}_{[i]} - \mathbf{Y}_{[i]}^*),$$

where

$$\begin{aligned} \bar{\mathbf{B}}_j^* &= \frac{\partial \bar{\mathbf{Y}}^*}{\partial \theta_j^*} (\Sigma^{\bar{\mathbf{Z}}})^{-1}, \\ \mathbf{B}_{i,j}^* &= \left(\frac{\partial \mathbf{Y}_{[i]}^*}{\partial \theta_j^*} \right)^T \mathbf{A}_i^T (\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}})^{-1} \mathbf{A}_i. \end{aligned}$$

We let $\boldsymbol{\xi}$ be the vector containing all the parameters in $\boldsymbol{\xi}_d$ as well as the emulator parameter being re-estimated. The partial derivative with respect to the k th parameter in $\boldsymbol{\xi}$, ξ_k , can be written as

$$\begin{aligned} \frac{\partial cl_n(\boldsymbol{\psi})}{\partial \xi_k} &\propto \frac{1}{2} (\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*)^T \bar{\mathbf{B}}_k^d (\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*) \\ &+ \frac{1}{2} \sum_{i=1}^M (\mathbf{Z}_{[i]} - \mathbf{Y}_{[i]}^*)^T \mathbf{B}_{i,k}^d (\mathbf{Z}_{[i]} - \mathbf{Y}_{[i]}^*) \\ &+ \sum_{i=1}^M (\mathbf{Z}_{[i]} - \mathbf{Y}_{[i]}^*)^T \tilde{\mathbf{B}}_{i,k}^d (\mathbf{Z}_{[i]} - \mathbf{Y}_{[i]}^*) \end{aligned}$$

where

$$\begin{aligned} \bar{\mathbf{B}}_k^d &= (\Sigma^{\bar{\mathbf{Z}}})^{-1} \frac{\partial \Sigma^{\bar{\mathbf{Z}}}}{\partial \xi_k} (\Sigma^{\bar{\mathbf{Z}}})^{-1}, \\ \mathbf{B}_{i,k}^d &= \mathbf{A}_i^T (\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}})^{-1} \frac{\partial \Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}}{\partial \xi_k} (\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}})^{-1} \mathbf{A}_i, \\ \tilde{\mathbf{B}}_{i,k}^d &= - \left(\frac{\partial \mathbf{A}_i}{\partial \xi_k} \right)^T (\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}})^{-1} \mathbf{A}_i. \end{aligned}$$

Inference on $\boldsymbol{\theta}^*$, our main goal, requires only calculating the asymptotic covariance of $\hat{\boldsymbol{\theta}}_n^B$ due to the asymptotic independence between $\hat{\boldsymbol{\theta}}_n^B$ and $\hat{\boldsymbol{\xi}}_n^B$, the posterior

modes of $\boldsymbol{\theta}^*$ and $\boldsymbol{\xi}$ respectively. More specifically, for any j and k ,

$$\text{Cov} \left(\frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial \theta_j^*}, \frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial \xi_k} \right) = 0,$$

because a linear combinations of zero-mean normal random variables and a quadratic form of the same variables are uncorrelated to one another. As a result, $\hat{\boldsymbol{\theta}}_n^B$ and $\hat{\boldsymbol{\xi}}_n^B$ have zero cross-covariance in \mathbf{G}_n and are asymptotically independent due to normality. Let \mathbf{P}_n^* be the part of \mathbf{P}_n , the covariance matrix between partial derivatives with respect to the parameters in $\boldsymbol{\theta}^*$ only. Likewise, let \mathbf{Q}_n^* be the part of \mathbf{Q}_n that contains only the negative expected Hessian of the parameters in $\boldsymbol{\theta}^*$. For inference on $\boldsymbol{\theta}^*$, it is sufficient to compute \mathbf{P}_n^* and \mathbf{Q}_n^* instead of \mathbf{P}_n and \mathbf{Q}_n .

We compute the (k, l) th element of \mathbf{P}_n^* by plugging in $\hat{\boldsymbol{\psi}}_n^B$ in place of $\boldsymbol{\psi}$ in

$$\begin{aligned} \text{Cov} \left(\frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial \theta_k^*}, \frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial \theta_l^*} \right) = & \quad \bar{\mathbf{B}}_k^* \bar{\boldsymbol{\Sigma}}^{\bar{\mathbf{Z}}} (\bar{\mathbf{B}}_l^*)^T \\ & + \sum_{i=1}^M \sum_{j=1}^M \mathbf{B}_{i,k}^* \boldsymbol{\Sigma}_{i,j}^{\mathbf{Z}} (\mathbf{B}_{j,l}^*)^T \\ & + \sum_{i=1}^M \bar{\mathbf{B}}_k^* \bar{\boldsymbol{\Sigma}}_i^{\bar{\mathbf{Z}}, \mathbf{Z}} (\mathbf{B}_{i,l}^*)^T \\ & + \sum_{i=1}^M \bar{\mathbf{B}}_l^* \bar{\boldsymbol{\Sigma}}_i^{\bar{\mathbf{Z}}, \mathbf{Z}} (\mathbf{B}_{i,k}^*)^T, \end{aligned}$$

where $\boldsymbol{\Sigma}_{i,j}^{\mathbf{Z}}$ is the $n_i \times n_j$ covariance matrix between $\mathbf{Z}_{[i]}$ and $\mathbf{Z}_{[j]}$, and $\bar{\boldsymbol{\Sigma}}_i^{\bar{\mathbf{Z}}, \mathbf{Z}}$ is the $1 \times n_i$ covariance matrix between $\bar{\mathbf{Z}}$ and $\mathbf{Z}_{[i]}$ under the probability model in (3). Similarly, the second order partial derivative of $c\ell_n(\boldsymbol{\psi})$ with respect to θ_j^* and θ_k^* is given by

$$\begin{aligned} \frac{\partial^2 c\ell_n(\boldsymbol{\psi})}{\partial \theta_j^* \partial \theta_k^*} \propto & \quad \left(\frac{\partial^2 \bar{\mathbf{Y}}^*}{\partial \theta_j^* \partial \theta_k^*} \right)^T (\boldsymbol{\Sigma}^{\bar{\mathbf{Z}}})^{-1} (\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*) \\ & - \left(\frac{\partial \bar{\mathbf{Y}}^*}{\partial \theta_j^*} \right)^T (\boldsymbol{\Sigma}^{\bar{\mathbf{Z}}})^{-1} \frac{\partial \bar{\mathbf{Y}}^*}{\partial \theta_k^*} \\ & + \sum_{i=1}^M \left(\frac{\partial^2 \mathbf{Y}_{[i]}^*}{\partial \theta_j^* \partial \theta_k^*} \right)^T \mathbf{B}_i^T (\boldsymbol{\Sigma}_i^{\mathbf{Z}|\bar{\mathbf{Z}}})^{-1} \mathbf{B}_i (\mathbf{Z}_{[i]} - \mathbf{Y}_{[i]}^*) \end{aligned}$$

$$- \sum_{i=1}^M \left(\frac{\partial \mathbf{Y}_{[i]}^*}{\partial \theta_j^*} \right)^T \mathbf{B}_i^T \left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}} \right)^{-1} \mathbf{B}_i \frac{\partial \mathbf{Y}_{[i]}^*}{\partial \theta_k^*}.$$

The (j, k) th element of \mathbf{Q}_n^* is computed by substituting $\boldsymbol{\psi}$ with $\hat{\boldsymbol{\psi}}_n^B$ in

$$\begin{aligned} -E \left(\frac{\partial \text{cl}_n(\boldsymbol{\psi})}{\partial \theta_j^* \partial \theta_k^*} \right) &\propto \left(\frac{\partial \bar{\mathbf{Y}}^*}{\partial \theta_j^*} \right)^T \left(\Sigma^{\bar{\mathbf{Z}}} \right)^{-1} \frac{\partial \bar{\mathbf{Y}}^*}{\partial \theta_k^*} \\ &+ \sum_{i=1}^M \left(\frac{\partial \mathbf{Y}_{[i]}^*}{\partial \theta_j^*} \right)^T \mathbf{B}_i^T \left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}} \right)^{-1} \mathbf{B}_i \frac{\partial \mathbf{Y}_{[i]}^*}{\partial \theta_k^*}. \end{aligned}$$

Computing \mathbf{P}_n^* and \mathbf{Q}_n^* requires finding the first-order derivatives of $\mathbf{Y}_{[1]}^*, \dots, \mathbf{Y}_{[M]}^*$, and $\bar{\mathbf{Y}}^*$. Since they are unknown functions of $\boldsymbol{\theta}^*$, we approximate them using the corresponding derivatives of the emulator output. The approximated derivatives of $\bar{\mathbf{Y}}^*$ and $\mathbf{Y}_{[i]}^*$ with respect to θ_j^* are given by

$$\begin{aligned} \frac{\partial \bar{\mathbf{Y}}^*}{\partial \theta_j^*} &= \left(\mathbf{I}_M \otimes \left(\frac{\partial \Sigma_{\boldsymbol{\theta}^* \boldsymbol{\theta}}}{\partial \theta_j^*} \Sigma_{\boldsymbol{\theta}}^{-1} \right) \right) \bar{\mathbf{Y}}, \\ \frac{\partial \mathbf{Y}_{[i]}^*}{\partial \theta_j^*} &= \left(\mathbf{I}_{n_i} \otimes \left(\frac{\partial \Sigma_{\boldsymbol{\theta}^* \boldsymbol{\theta}}}{\partial \theta_j^*} \Sigma_{\boldsymbol{\theta}}^{-1} \right) \right) \mathbf{Y}_{[i]}. \end{aligned}$$

The derivative term $\frac{\partial \Sigma_{\boldsymbol{\theta}^* \boldsymbol{\theta}}}{\partial \theta_j^*}$ is determined by the covariance function for the parameter space. For the exponential covariance function used in our example, the derivative is

$$\left\{ \frac{\partial \Sigma_{\boldsymbol{\theta}^* \boldsymbol{\theta}}}{\partial \theta_i^*} \right\}_j = \phi_{\boldsymbol{\theta}, i} (-1)^{1(\theta_i^* > \theta_{ij})} \exp \left(- \sum_{k=1}^q \phi_{\boldsymbol{\theta}, k} |\theta_k^* - \theta_{kj}| \right),$$

for $i = 1, \dots, q$ and $j = 1, \dots, p$ where $1(\cdot)$ is the indicator function, and θ_{ij} is the i th parameter value of the j th design point $\boldsymbol{\theta}_j$.

Bibliography

- ADRAGNI, K. P. and COOK, R. D. (2009). Sufficient dimension reduction and prediction in regression. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **367** 4385–4405.
- ALEXANDER, L. V., ALLEN, S. K., BINDOFF, N. L., BRON, F. M., CHURCH, J. A., CUBASCH, U., EMORI, S., FORSTER, P., FRIEDLINGSTEIN, P., GILLETT, N., GREGORY, J. M., HARTMANN, D. L., JANSEN, E., KIRTMAN, B., KNUTTI, R., KANIKICHARLA, K. K., LEMKE, P., MAROTZKE, J., MASSON-DELMOTTE, V., MEEHL, G. A., MOKHOV, I. I., PIAO, S., PLATTNER, G. K., DAHE, Q., RAMASWAMY, V., RANDALL, D., RHEIN, M., ROJAS, M., SABINE, C., SHINDELL, D., STOCKER, T. F., TALLEY, L. D., VAUGHAN, D. G. and XIE, S. P. (2013). *Climate Change 2013: The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Cambridge University Press, Cambridge.
- ALLEY, R., BERNTSEN, T., BINDOFF, N. L., CHEN, Z., CHIDTHAISONG, A., FRIEDLINGSTEIN, P., GREGORY, J., HEGERL, G., HEIMANN, M., HEWITSON, B., JOOS, F., JOUZEL, J., KATTSOV, V., LOHMANN, U., MANNING, M., MATSUNO, T., MOLINA, M., NICHOLLS, N., OVERPECK, J., QIN, D., RAGA, G., RAMASWAMY, V., REN, J., RUSTICUCCI, M., SOLOMON, S., SOMERVILLE, R., STOCKER, T., STOTT, P., WHETTON, P., WOOD, R. and WRATT, D. (2007). *Climate Change 2007: The Physical Science Basis: Summary for Policymakers: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge.
- ALLEY, R. B., CLARK, P. U., HUYBRECHTS, P. and JOUGHIN, I. (2005). Ice-sheet and sea-level changes. *Science*, **310** 456–460.
- AMEMIYA, T. (1985). *Advanced econometrics*. Harvard University Press, New York.

- ANDRONOVA, N., SCHLESINGER, M., DESSAI, S., HULME, M. and LI, B. (2007). The concept of climate sensitivity: history and development. In *Human-induced Climate Change: An Interdisciplinary Assessment* (M. Schlesinger, H. Kheshgi, J. Smith, F. de la Chesnaye, J. M. Reilly, T. Wilson and C. Kolstad, eds.). Cambridge University Press, New York, 11–17.
- ANTONOV, J., SEIDOV, D., BOYER, T., LOCARNINI, R., MISHONOV, A., GARCIA, H., BARANOVA, O., ZWENG, M. and JOHNSON, D. (2010). *World Ocean Atlas 2009, Volume 2: Salinity*, S. Levitus, Ed., NOAA Atlas NESDIS 69, US Government Printing Office, Washington, DC 184.
- APPLEGATE, P., KIRCHNER, N., STONE, E., KELLER, K. and GREVE, R. (2012). An assessment of key model parametric uncertainties in projections of Greenland Ice Sheet behavior. *The Cryosphere*, **6** 589–606.
- ARENDRT, P. D., APLEY, D. W. and CHEN, W. (2012). Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *Journal of Mechanical Design*, **134** 100908.
- ARTEMIUO, A. and LI, B. (2009). On principal components and regression: A statistical explanation of a natural phenomenon. *Statistica Sinica*, **19** 1557–1565.
- BAGLAMA, J. and REICHEL, L. (2005). Augmented implicitly restarted lanczos bidiagonalization methods. *SIAM Journal on Scientific Computing*, **27** 19–42.
- BAMBER, J., LAYBERRY, R. and GOGINENI, S. (2001). A new ice thickness and bed data set for the greenland ice sheet: 1. measurement, data reduction, and errors. *Journal of Geophysical Research: Earth Surface*, **106** 33773–33780.
- BAMBER, J. L. and ASPINALL, W. (2013). An expert judgement assessment of future sea level rise from the ice sheets. *Nature Climate Change*, **3** 424–427.
- BANERJEE, A., DUNSON, D. and TOKDAR, S. (2013). Efficient Gaussian Process Regression for Large Data Sets. *Biometrika*, **100** 75–89.
- BANERJEE, S., GELFAND, A., FINLEY, A. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **70** 825–848.
- BARNES, S. L. (1964). A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, **3** 396–409.
- BASTOS, L. S. and O’HAGAN, A. (2009). Diagnostics for gaussian process emulators. *Technometrics*, **51** 425–438.

- BAYARRI, M., BERGER, J., CAFEO, J., GARCIA-DONATO, G., LIU, F., PALOMO, J., PARTHASARATHY, R., PAULO, R., SACKS, J. and WALSH, D. (2007). Computer model validation with functional output. *The Annals of Statistics*, **35** 1874–1906.
- BEAUMONT, M., ZHANG, W. and BALDING, D. (2002). Approximate Bayesian computation in population genetics. *Genetics*, **162** 2025–2035.
- BERGER, J., DE OLIVEIRA, V. and SANSÓ, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, **96** 1361–1374.
- BESAG, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, **24** 179–195.
- BESAG, J. (1977). Efficiency of pseudolikelihood estimation for simple gaussian fields. *Biometrika*, **64** 616–618.
- BHAT, K., HARAN, M. and GOES, M. (2010). Computer model calibration with multivariate spatial output: A case study. In *Frontiers of Statistical Decision Making and Bayesian Analysis* (M. Chen, P. Müller, D. Sun, K. Ye and D. Dey, eds.). Springer, 168–184.
- BHAT, K., HARAN, M., OLSON, R. and KELLER, K. (2012). Inferring likelihoods and climate system characteristics from climate models and multiple tracers. *Environmetrics*, **23** 345–362.
- BRYAN, F. (1987). Parameter sensitivity of primitive equation ocean general circulation models. *Journal of Physical Oceanography*, **17** 970–985.
- BRYDEN, H. L. (1973). New polynomials for thermal expansion, adiabatic temperature gradient and potential temperature of sea water. *Deep Sea Research and Oceanographic Abstracts*, **20** 401–408.
- CARAGEA, P. and SMITH, R. (2006). Approximate likelihoods for spatial processes. *Preprint*.
- CASTRUCCIO, S., MCINERNEY, D. J., STEIN, M. L., LIU, F., JACOB, R. L. and MOYER, E. J. (2013a). Statistical emulation of climate model projections based on precomputed gcm runs. *Journal of Climate*, **27**.
- CASTRUCCIO, S., STEIN, M. L. ET AL. (2013b). Global space–time models for climate ensembles. *The Annals of Applied Statistics*, **7** 1593–1611.

- CHANG, W., APPLGATE, P., HARAN, H. and KELLER, K. (2013a). Probabilistic calibration of a greenland ice sheet model using spatially-resolved synthetic observations: toward projections of ice mass loss with uncertainties. *submitted to Geoscientific Model Development Discussions*.
- CHANG, W., HARAN, M., OLSON, R. and KELLER, K. (2013b). A composite likelihood approach to computer model calibration using high-dimensional spatial data. *arXiv preprint arXiv:1308.0049, Statistica Sinica, in press*.
- CHANG, W., HARAN, M., OLSON, R. and KELLER, K. (2014). Fast dimension-reduced climate model calibration and the effect of data aggregation. *The Annals of Applied Statistics*, **8** 649–673.
- CHERNOZHUKOV, V. and HONG, H. (2003). An mcmc approach to classical estimation. *Journal of Econometrics*, **115** 293–346.
- COLLINS, M., DASGUPTA, S. and SCHAPIRE, R. E. (2002). A generalization of principal components analysis to the exponential family. In *Advances in Neural Information Processing Systems 14* (T. Dietterich, S. Becker and Z. Ghahramani, eds.). MIT Press, 617–624.
- COOK, R. (1998). *Regression Graphics: Ideas for Studying Regressions Through Graphics*. Wiley. New York.
- COOK, R. D. (2004). Testing predictor contributions in sufficient dimension reduction. *The Annals of Statistics*, **32** 1062–1092.
- COX, D. and REID, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91** 729–737.
- CRESSIE, N. (1993). *Statistics for Spatial Data*. Wiley, New York.
- CRESSIE, N. (1994). Models for spatial processes. *Statistical Methods for Physical Science*, **28** 93–124.
- CRESSIE, N. and HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94** 1330–1339.
- 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** 209–226.
- CURRIERO, F. C. and LELE, S. (1999). A composite likelihood approach to semi-variogram estimation. *Journal of Agricultural, Biological, and Environmental Statistics*, **4** 9–28.

- DAVID, M. (1990). Usage summary for selected optimization routines. *Technical Report*. AT&T Bell Laboratories Technology Computing Science Technical Report No. 153.
- DE OLIVEIRA, V. (2005). Bayesian inference and prediction of gaussian random fields based on censored data. *Journal of Computational and Graphical Statistics*, **14**.
- DIEBOLT, J. and IP, E. (1996). Stochastic EM: method and application. In *Markov Chain Monte Carlo in Practice* (W. Gilks, S. Richardson and D. Spiegelhalter, eds.). Springer, 259–273.
- DRIGNEI, D., FOREST, C. and NYCHKA, D. (2008). Parameter estimation for computationally intensive nonlinear regression with an application to climate modeling. *The Annals of Applied Statistics*, **2** 1217–1230.
- EIDSVIK, J., SHABY, B., REICH, B., WHEELER, M. and NIEMI, J. (2013). Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics* doi:10.1080/10618600.2012.760460.
- FLEGAL, J. M., HARAN, M. and JONES, G. L. (2008). Markov chain monte carlo: Can we trust the third significant figure? *Statistical Science*, **23** 250–260.
- FOFONOFF, N. (1977). Computation of potential temperature of seawater for an arbitrary reference pressure. *Deep-Sea Research*, **24** 489–491.
- FOREST, C., STONE, P. and SOKOLOV, A. (2008). Constraining climate model parameters from observed 20th century changes. *Tellus A*, **60** 911–920.
- FOREST, C. E., STONE, P. H., SOKOLOV, A. P., ALLEN, M. R. and WEBSTER, M. D. (2002). Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295** 113–117.
- FUENTES, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, **102** 321–331.
- FURRER, R., GENTON, M. G. and NYCHKA, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, **15** 502–523.
- GAY, D. M. (1983). Algorithm 611: Subroutines for unconstrained minimization using a model/trust-region approach. *ACM Transactions on Mathematical Software*, **9** 503–524.
- GELFAND, A. E., DIGGLE, P., GUTTORP, P. and FUENTES, M. (2010). *Handbook of Spatial Statistics*. CRC Press. New York.

- GLADSTONE, R. M., LEE, V., ROUGIER, J., PAYNE, A. J., HELLMER, H., LE BROCCQ, A., SHEPHERD, A., EDWARDS, T. L., GREGORY, J. and CORNFORD, S. L. (2012). Calibrated prediction of pine island glacier retreat during the 21st and 22nd centuries with a coupled flowline model. *Earth and Planetary Science Letters*, **333** 191–199.
- GNANADESIKAN, A. (1999). A simple predictive model for the structure of the oceanic pycnocline. *Science*, **283** 2077–2079.
- GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, **97** 590–600.
- GNEITING, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, **19** 1327–1349.
- GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **69** 243–268.
- GODAMBE, V. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, **31** 1208–1211.
- GOES, M., URBAN, N., TONKONOJENKOV, R., HARAN, M., SCHMITTNER, A. and KELLER, K. (2010). What is the skill of ocean tracers in reducing uncertainties about ocean diapycnal mixing and projections of the Atlantic Meridional Overturning Circulation? *Journal of Geophysical Research: Oceans*, **115** C10042, doi:10.1029/2010JC006106.
- GOURIEROUX, C., MONFORT, A. and RENAULT, E. (2008). Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, **103** 1119–1130.
- 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** 901–918.
- GREVE, R., OTSU, S. ET AL. (2007). The effect of the north-east ice stream on the greenland ice sheet in changing climates. *The Cryosphere Discussions*, **1** 41–76.
- GREVE, R., SAITO, F. and ABE-OUCHI, A. (2011). Initial results of the searise numerical experiments with the models sicopolis and icies for the greenland ice sheet. *Annals of Glaciology*, **52** 23–30.
- GROVES, D. G. and LEMPert, R. J. (2007). A new analytic method for finding policy-relevant scenarios. *Global Environmental Change*, **17** 73–85.

- HALKO, N., MARTINSSON, P.-G. and TROPP, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, **53** 217–288.
- HANSEN, J., RUSSELL, G., LACIS, A., FUNG, I., RIND, D. and STONE, P. (1985). Climate response times - dependence on climate sensitivity and ocean mixing. *Science*, **229** 857–859.
- HARVILLE, D. (2008). *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, Berlin.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.
- HEAGERTY, P. J. and LELE, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, **93** 1099–1111.
- HIGDON, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, **5** 173–190.
- HIGDON, D. (2002). Space and space-time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues* (C. Anderson, V. Barnett, P. Chatwin and A. El-Shaarawi, eds.). Springer, 37–56.
- HIGDON, D., GATTIKER, J., WILLIAMS, B. and RIGHTLEY, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, **103** 570–583.
- HIGDON, D., REESE, C., MOULTON, J., VRUGT, J. and FOX, C. (2009). Posterior exploration for computationally intensive forward models. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X. Meng, eds.). CRC Press, New York, 401–418.
- HIGDON, D., SWALL, J. and KERN, J. (1999). Non-stationary spatial modeling. *Bayesian Statistics*, **6** 761–768.
- HUNTER, D. R. and LANGE, K. (2004). A tutorial on mm algorithms. *The American Statistician*, **58** 30–37.
- JONES, G., 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** 1537–1547.
- JORDAN, K. (1961). Discrete representations of random signals. *Technical Report*. Massachusetts Institute of Technology Technical Report 378.

- KAUFMAN, C. G., SCHERVISH, M. J. and NYCHKA, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, **103** 1545–1555.
- KELLER, K., BOLKER, B. and BRADFORD, D. F. (2004). Uncertain climate thresholds and optimal economic growth. *Journal of Environmental Economics and Management*, **48** 723–741.
- KELLER, K., DEUTSCH, C., HALL, M. G. and BRADFORD, D. F. (2007). Early detection of changes in the north atlantic meridional overturning circulation: implications for the design of ocean observation systems. *Journal of Climate*, **20** 145–157.
- KELLER, K., HALL, M., KIM, S.-R., BRADFORD, D. F. and OPPENHEIMER, M. (2005). Avoiding dangerous anthropogenic interference with the climate system. *Climatic Change*, **73** 227–238.
- KENNEDY, M. and O’HAGAN, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **63** 425–464.
- KEY, R., KOZYR, A., SABINE, C., LEE, K., WANNINKHOF, R., BULLISTER, J., FEELY, R., MILLERO, F., MORDY, C. and PENG, T. (2004). A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global Biogeochemical Cycles*, **18**.
- KNUTTI, R. and HEGERL, G. C. (2008). The equilibrium sensitivity of the Earth’s temperature to radiation changes. *Nature Geoscience*, **1** 735–743.
- KNUTTI, R., STOCKER, T. F., JOOS, F. and PLATTNER, G. K. (2002). Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature*, **416** 719–723.
- KRIEGLER, E. (2005). *Imprecise probability analysis for integrated assessment of climate change*. Ph.D. thesis, Universitätsbibliothek.
- KRIGE, D. G. (1951). A Statistical approach to some mine valuation and allied problems on the witwatersrand. *Journal of the Chemical Metallurgical and Mining Society of South Africa*, **52** 119–139.
- LANGE, K., HUNTER, D. R. and YANG, I. (2000). Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, **9** 1–20.
- LEE, S., HUANG, J. Z. and HU, J. (2010). Sparse logistic principal components analysis for binary data. *The Annals of Applied Statistics*, **4** 1579–1601.

- LEMKE, P., REN, J., ALLEY, R. B., ALLISON, I., CARRASCO, J., FLATO, G., FUJII, Y., KASER, G., MOTE, P. and THOMAS, R. H. (2007). *Climate Change 2007: The Physical Science Basis: Summary for Policymakers: Technical Summary and Frequently Asked Questions: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Cambridge University Press, Cambridge.
- LEVITUS, S., ANTONOV, J., BOYER, T., BARANOVA, O., GARCIA, H., LOCARNINI, R., MISHONOV, A., REAGAN, J., SEIDOV, D., YAROSH, E. ET AL. (2012). World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophysical Research Letters*, **39**.
- LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86** 316–327.
- 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** 423–498.
- LINDSAY, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, **80** 221–39.
- LINDSAY, B. G., YI, G. Y. and SUN, J. (2011). Issues and strategies in the selection of composite likelihoods. *Statistica Sinica*, **21** 71.
- LITTLE, C. M., OPPENHEIMER, M., ALLEY, R. B., BALAJI, V., CLARKE, G. K. C., DELWORTH, T. L., HALLBERG, R., HOLLAND, D. M., HULBE, C. L., JACOBS, S., JOHNSON, J. V., LEVY6WILLIAM H. LIPSCOMB11SHAWN J. MARSHALL12BYRON R. PARIZEK13ANTONY J. PAYNE14GAVIN A. SCHMIDT, H., STOUFFER, R. J., VAUGHAN, D. G. and WINTON, M. (2007). Toward a new generation of ice sheet models. *Eos, Transactions American Geophysical Union*, **88** 578–579.
- LIU, F., BAYARRI, M. and BERGER, J. (2009). Modularization in bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, **4** 119–150.
- LOCARNINI, R., MISHONOV, A., ANTONOV, J., BOYER, T., GARCIA, H., BARANOVA, O., ZWENG, M. and JOHNSON, D. (2010). *World Ocean Atlas 2009, Volume 1: Temperature*, S. Levitus, Ed., NOAA Atlas NESDIS 68, US Government Printing Office, Washington, DC 184.
- LOVETT, J. R. (1978). Merged seawater sound-speed equations. *Journal of the American Statistical Association*, **63** 1713–1718.

- MARDIA, K. and MARSHALL, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, **71** 135–146.
- MATÉRN, B. (1960). Spatial variation. Meddelanden fran Statens Skogsforsknings Institute, Stockholm. Band 49, No. 5.
- MATHÉRON, G. (1963). Principles of geostatistics. *Economic geology*, **58** 1246–1266.
- MCNEALL, D., CHALLENGOR, P., GATTIKER, J. and STONE, E. (2013). The potential of an observational data set for calibration of a computationally expensive computer model. *Geoscientific Model Development*, **6** 1715–1728.
- MOLENBERGHS, G. and VERBEKE, G. (2005). *Models for Discrete Longitudinal Data*. Springer, New York.
- NIELSEN, S. F. (2000). The stochastic em algorithm: estimation and asymptotic results. *Bernoulli* 457–489.
- NORDHAUS, W. D. and BOYER, J. (2000). *Warning the World: Economic Models of Global Warming*. MIT Press, Cambridge.
- NOWICKI, S., BINDSCHADLER, R. A., ABE-OUCHI, A., ASCHWANDEN, A., BUELER, E., CHOI, H., FASTOOK, J., GRANZOW, G., GREVE, R., GUTOWSKI, G. ET AL. (2013). Insights into spatial sensitivities of ice mass response to environmental change from the searise ice sheet modeling project ii: Greenland. *Journal of Geophysical Research: Earth Surface*, **118** 1025–1044.
- NYCHKA, D., BANDYOPADHYAY, S., HAMMERLING, D., LINDGREN, F. and SAIN, S. (2013). A multi-resolution gaussian process model for the analysis of large spatial data sets. *NCAR Technical Note NCAR/TN-504+STR*. Doi: 10.5065/D6VH5KTJ.
- OLSON, R., SRIVER, R., GOES, M., URBAN, N., MATTHEWS, H., HARAN, M. and KELLER, K. (2012). A climate sensitivity estimate using Bayesian fusion of instrumental observations and an Earth System model. *Journal of Geophysical Research: Atmospheres*, **117** D04103, doi:10.1029/2011JD016620.
- OLSON, R., SRIVER, R., HARAN, M., CHANG, W., URBAN, N. and KELLER, K. (2013). What is the effect of unresolved internal climate variability on climate sensitivity estimates? *Journal of Geophysical Research: Atmospheres*, **118** 4348–4358.
- O'REILLY, J., ORESKES, N. and OPPENHEIMER, M. (2012). The rapid disintegration of projections: the west antarctic ice sheet and the intergovernmental panel on climate change. *Social studies of science*, **42** 709–731.

- PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, **105** 263–277.
- RAHMSTORF, S. (1997). Risk of sea-change in the atlantic. *Nature*, **388** 825–826.
- RIBATET, M., COOLEY, D. and DAVISON, A. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, **22** 813–846.
- ROUGIER, J. (2008). Comment on article by Sansó et al. *Bayesian Analysis*, **3** 45–56.
- SACKS, J., WELCH, W., MITCHELL, T. and WYNN, H. (1989). Design and analysis of computer experiments. *Statistical Science*, **4** 409–423.
- SANSÓ, B. and FOREST, C. (2009). Uncertainty quantification: Statistical calibration of climate system properties. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **58** 485–03.
- SCHMITTNER, A., URBAN, N., KELLER, K. and MATTHEWS, D. (2009). Using tracer observations to reduce the uncertainty of ocean diapycnal mixing and climate–carbon cycle projections. *Global Biogeochemical Cycles*, **23** 737–750.
- SCHNEIDER, S. H. (2001). What is ‘dangerous’ climate change? *Nature*, **411** 17–19.
- SHABY, B. (2014). The open-faced sandwich adjustment for mcmc using estimating functions. *Journal of Computational and Graphical Statistics*, **23** 853–876.
- SHABY, B. and RUPPERT, D. (2012). Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics*, **21** 433–452.
- SIMPSON, D., LINDGREN, F. and RUE, H. (2012). In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics*, **23** 65–74.
- SRIVER, L., NATHAN, M., OLSON, R. and KELLER, K. (2012). Toward a physically plausible upper bound of sea-level rise projections. *Climatic Change*, **115** 893–902.
- STEIN, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag, Berlin.
- STEIN, M. (2014). Limitations on low rank approximations for covariance matrices of spatial data. *Spatial Statistics*, **8** 1–19.

- STEIN, M. L. (2005). Space–time covariance functions. *Journal of the American Statistical Association*, **100** 310–321.
- STEIN, M. L., CHI, Z. and WELTY, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **66** 275–296.
- SWEETING, T. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *The Annals of Statistics*, **8** 1375–1381.
- TAVARÉ, S., BALDING, D. J., GRIFFITHS, R. C. and DONNELLY, P. (1997). Inferring coalescence times from dna sequence data. *Genetics*, **145** 505–518.
- TIERNEY, L., ROSSINI, A. J., LI, N. and SEVCIKOVA, H. (2013). *snow: Simple Network of Workstations*. R package version 0.3-3.
- UNESCO (1981). Tenth report of the joint panel on oceanographic tables and standards. *Technical Report*. UNESCO Technical Reports on Marine Science 36.
- URBAN, N. and KELLER, K. (2010). Probabilistic hindcasts and projections of the coupled climate, carbon cycle and Atlantic meridional overturning circulation system: a Bayesian fusion of century-scale observations with a simple model. *Tellus A*, **62** 737–750.
- VARIN, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, **92** 1–28.
- VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **50** 297–312.
- WALKER, W. E., HARREMOËS, P., ROTMANS, J., VAN DER SLUIJS, J. P., VAN ASSELT, M. B., JANSSEN, P. and KRAYER VON KRAUSS, M. P. (2003). Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, **4** 5–17.
- WEAVER, A., EBY, M., WIEBE, E., BITZ, C., DUFFY, P., EWEN, T., FANNING, A., HOLLAND, M., MACFADYEN, A. and MATTHEWS, H. (2001). The UVic Earth System Climate Model: Model description, climatology, and applications to past, present and future climates. *Atmosphere-Ocean*, **39** 361–428.
- WHITE, H. (1996). *Estimation, Inference and Specification Analysis*. Cambridge University Press, Cambridge.
- WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika*, **41** 434–449.

- WOODBURY, M. (1950). Inverting modified matrices. *Memorandum Report, Statistical Research Group, Princeton University*, **42** 106.
- WUNSCH, C. and FERRARI, R. (2004). Vertical mixing, energy, and the general circulation of the oceans. *Annual Review of Fluid Mechanics*, **36** 281–314.
- WYNN, H. (2001). Contribution to the discussion on the paper by Kennedy and O’Hagan. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **63** 425–464.
- YAKOWITZ, S. and SZIDAROVSKY, F. (1985). A comparison of kriging with nonparametric regression methods. *Journal of Multivariate Analysis*, **16** 21–53.
- ZHANG, H. and ZIMMERMAN, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, **92** 921–936.
- ZHU, Z. and WU, Y. (2010). Estimation and prediction of a class of convolution-based spatial nonstationary models for large spatial data. *Journal of Computational and Graphical Statistics*, **19** 74–95.

Vita

Won Chang

Education

- Ph.D. Statistics, Pennsylvania State University, 2014.
Thesis title: *Climate model calibration using high-dimensional and non-Gaussian spatial data* (Thesis advisor: Dr. Murali Haran, Thesis co-advisor: Dr. Klaus Keller).
- M.S. Statistics, Korea University, 2009.
Thesis title: *Estimating volatility and distribution of European option prices using Bayesian UHF GARCH-M model* (Thesis advisor: Dr. Yousung Park).
- B.S. Statistics, Korea University, 2007.

Publications

- Chang, W., M. Haran, R. Olson, and K. Keller (2014) Fast dimension-reduced climate model calibration, *the Annals of Applied Statistics*, 8 (2), 649-673.
- Chang, W., M. Haran, R. Olson, K. Keller (2014) A composite likelihood approach to computer model calibration with high-dimensional spatial data, *Statistica Sinica*, *in press*.
- Chang, W., P. J. Applegate, M. Haran, K. Keller (2014) Computer model calibration using high-dimensional spatial binary data, *in preparation*.
- Chang, W., P. J. Applegate, M. Haran, K. Keller (2014) Probabilistic calibration of a Greenland Ice Sheet model using spatially-resolved synthetic observations: toward projections of ice mass loss with uncertainties, *submitted to Geoscientific Model Development Discussions*.
- Olson, R., R. Srivier, W. Chang, M. Haran, N. M. Urban and K. Keller (2013) What is the effect of unresolved internal climate variability on climate sensitivity estimates? *Journal of Geophysical Research - Atmospheres*, 118 (10), 4348–4358.

Honors and Awards

- Winner, 2014 American Statistical Association Section on Statistics and the Environment Student Paper Competition.
- Graduate Fellow, Eberly College of Science, The Pennsylvania State University, Fall 2009 – Spring 2010.