STATISTICAL EMULATION AND UNCERTAINTY QUANTIFICATION IN COMPUTER EXPERIMENTS

BY

LINGLIN HE

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics and Biostatistics

Written under the direction of

Ying Hung

and approved by

New Brunswick, New Jersey

May, 2019

ABSTRACT OF THE DISSERTATION

Statistical Emulation and Uncertainty Quantification in Computer Experiments

by Linglin He

Dissertation Director: Ying Hung

Computer experiments refer to the study of real systems using complex mathematical models. They have been widely used as alternatives to physical experiments, especially for studying complex systems in science and engineering. Typically, computer experiments require a great deal of time and computing to conduct and they are nearly deterministic in the sense that a particular input will produce almost the same output if given to the computer experiment on another occasion. Therefore, it is desirable to build an interpolator for computer experiment outputs and use it as an emulator for the actual computer experiment. This thesis mainly focuses on building efficient statistical emulators for computer experiments and providing prediction uncertainty with real-world applications.

Gaussian process (GP) models are widely used in the analysis of computer experiments. However, two issues have not been solved satisfactorily. One is the computational issue that hinders GP from broader application, especially for massive data with

high-dimensional inputs. The other is the underestimation of prediction uncertainty in GP modeling. To tackle these problems simultaneously, in Chapter 1 we propose two methods to construct GP predictive distributions based on a new version of bootstrap subsampling. The new subsampling procedure borrows the strength of space-filling designs to provide an efficient subsample and thus reduce the computational complexity. It is shown that this procedure not only alleviates the computational difficulty in GP modeling, but also provides unbiased predictors with better quantifications of uncertainty comparing with existing methods. We illustrate the proposed methods by two complex computer experiments with high-dimensional inputs and tens of thousands of simulation outputs.

Traditional GP models are limited in the computational capability of dealing with massive and complex data. To overcome the computational issue without imposing strong assumptions, a spline-based approach is developed to build emulators for computer experiments to handle big spatial-temporal data in Chapter 2. A direct application of the proposed framework is to model Antarctic ice-sheet contributions to sea level rise. Sea level rise is expected to impact millions of people in coastal communities in the coming centuries. Global, regional, and local sea level rise projections are highly uncertain, partially due to uncertainties in Antarctic ice-sheet (AIS) dynamics, and parameterized simulations are expensive to run. We create an ice-sheet emulator to provide near-continuous distributions of the sea-level equivalent of AIS melt based on two input parameters, which alter the behavior AIS mass loss, under two forcing scenarios: the Last Interglacial and RCP 8.5 forcing. The spline-based emulator with Gaussian Process priors on the spline parameters is flexible enough to capture the nonlinearity of the underlying structure, while computationally feasible at the same time. It achieves a good fit for the physical model and provides prediction uncertainties simultaneously.

The reminder of this thesis is organized as follows. In Chapter 1, we introduce two methods to construct GP predictive distribution using design-based subsampling. In Chapter 2, a spline-based approach with GP priors in parameters is proposed for emulating Antarctic ice-sheet contributions to sea level rise.

Acknowledgements

First I would like to express my deepest appreciation to my thesis advisor, Professor Ying Hung. Without her ongoing guidance, brilliant ideas, strong encouragement and support, this dissertation would not have been accomplished. More than an advisor in my research, she is also a mentor and a friend of mine, doing everything in her power to instruct me towards a successful future career. Her professional knowledge, continuous pursuit on advances in statistical research and devotion to students inspire me a lot, and stimulate me to go further. I am truly blessed having her as my advisor.

I also would like to express my gratitude to all my Ph.D. dissertation committee members. I especially thank Professor Tirthankar Dasgupta for his support and intellectual discussions on my research after he joined Rutgers University. I would like to extend my thanks to other committee members, Professor Han Xiao and Professor Myong K. Jeong for their constructive comments and feedback on my dissertation. Besides, I would also like to extend my thanks to all my Ph.D. qualifying oral examination comittee members, Professor William Edward Strawderman, Professor Dan Yang and Professor Cun-hui Zhang for their advices and insights on my research topics.

My thanks also go to the Department of Statistics of Rutgers University for providing me financial support and all its faculty members and staff for creating the great research environment. Thanks to the graduate directors Professor Harry Crane and Professor Tirthankar Dasgupta, and also to Professor John Kolassa, the former graduate director, for their valuable advices and kind support during my graduate study. Sincerely thanks to the department chair, Professor Regina Liu, for her unremitting efforts to make sure every student is taken care of. Special thanks to my friends and colleagues at Rutgers University, especially Xinyu Sun and Yilei Zhan, for sharing my happiness and accompanying me during my difficult time. I will always cherish these warm and sweet memories we experienced.

I would also like to take this opportunity to thank Professor Micheal Stein and Professor Robert Kopp for their valuable suggestions and insightful comments, as well as the opportunity to collaborate with the group from Department of Earth and Planetary Sciences on the project in Chapter 2, who are Erica L. Ashe, Daniel Gilford and Bryan Raney. My thanks especially go to Daniel Gilford, for proposing the project, his generosity and patience in sharing his knowledge in ice sheet model and sea level rise, which motivated and contributes to several aspects of the dissertation project in Chapter 2.

Finally, great thanks to my family for their unconditional love and my boyfriend for his company and encouragement.

Dedication

To my parents Jianzhong and Qiufang; To my boyfriend Liang

Table of Contents

List of Figures

List of Tables

Chapter 1

Efficient Gaussian Process Prediction using Design-Based Subsampling

1.1 Introduction

Computer experiments refer to the study of real systems using complex mathematical models. They have been widely used as alternatives to physical experiments, especially for studying complex systems. The reason is, in many situations, a physical experiment is infeasible because it is unethical, impossible, inconvenient or too expensive. A mathematical model of the system can often be developed and input/output pairs can be produced with the help of computers. Computer experiments are widely used in science and engineering. Typically, computer experiments require a great deal of time and computing to conduct and they are nearly deterministic in the sense that a particular input will produce almost the same output if given to the computer experiment on another occasion. Therefore, it is desirable to build an interpolator for computer experiment outputs and use it as an emulator for the actual computer experiment. More discussions of design and analysis of computer experiments can be found in Santner et al. (2003) and Fang et al. (2006).

A Gaussian process (GP) model (or called kriging) is a flexible and widely used method in the analysis of computer experiments; however, there are two critical issues in GP modeling. One is the computational issue that hinders GP from broader applications, especially with high-dimensional inputs and massive outputs observed on irregular grids. This is because modeling and prediction of GP heavily involve manipulations of the $N \times N$ correlation matrix, where N is the sample size, that require $O(N^3)$ computations and often result in singularity. This issues is even more critical for the analysis of complex computer experiments because the estimation of high-dimensional correlation parameters often leads to numerical instability in the estimation and prediction. The second issue is how to accurately quantify the uncertainty based on GP. It is well-known that the GP predictive interval constructed by substituting the true parameters by estimators, often called plug-in predictor, underestimates the uncertainty (Santner et al. 2003, p.98). Although there are intensive studies addressing both issues, to the best of our knowledge, there is no systematic approach that can address both issues simultaneously, which is the main focus of this paper.

The computational issue is well recognized in the literature and a number of methods have been proposed. Various methods address this problem by changing the model to one that is computationally convenient. Examples include Rue and Tjelmeland (2002), Rue and Held (2005), Cressie and Johannesson (2008), Banerjee et al. (2008), Gramacy and Lee (2008), Wikle (2010), Chang et al. (2014), Castrillon et al. (2015). Another approach is to approximate the likelihood for the original data. Examples include Nychka (2000), Smola and Bartlett (2001), Nychka et al. (2002), Stein et al. (2004), Furrer et al. (2006), Snelson and Ghahramani (2006), Fuentes (2007), Kaufman et al. (2008), Gramacy and Apley (2014), Nychka et al. (2015). Despite various methods, most of the existing ones are developed for data sets collected from a regular grid under a low-dimensional geostatistical setting. These assumptions are often violated in computer experiments because having high-dimensional inputs is common in complex computer experiments and the computational expense often prohibits running computer experiments over a dense grid of input configurations. A commonly used approach for high-dimensional computer experiments is to impose a sparsity constraint on the correlation matrix (Kaufman et al. 2008, 2011). However, it has been shown that this method does not work well for purposes of parameter estimation (Stein 2013, Liang et al. 2013), which is crucial for the GP predictor. In addition, the connection between the degree of sparsity and computation time is nontrivial.

The second issue of uncertainty quantification in GP prediction is important but has been overlooked in the literature. For example, most of the aforementioned methods address the computational issue, but adopt the idea of plugin predictors for inference and therefore underestimate the prediction uncertainty. Moreover, with different approximation techniques, these methods bring in additional uncertainty which is difficult to quantify. Although some methods, such as Bayesian approaches (Kennedy and O‘Hagan 2001, Schmidt and O‘Hagan 2003) and regular bootstrap (Santner et al. 2003, Luna and Young 2003), have been proposed to provide a better quantification of prediction uncertainty, they are computationally intensive and often intractable for massive data.

In this paper, a new framework is proposed to construct GP predictors and their predictive distributions. The proposed methods are easy to compute and more accurate in quantifying the predictive uncertainty comparing with existing methods. The idea is to combine the bootstrap predictive distribution with an experimental design-based stratified subsampling plan. Bootstrap is an increasingly utilized method for obtaining accurate confidence intervals and performing statistical inference (Efron 1979, DiCiccio and Efron 1992, 1996, Efron and Tibshirani 1993). Direct application of bootstrap methods to construct predictive distributions for GP is conceptually attractive but computationally prohibitive especially for massive data. Therefore, we introduce a new version of bootstrap using design-based subsampling and propose two methods for the construction of bootstrap predictive distributions. It can be shown that the proposed predictors are unbiased, given a significant alleviation in computation. Moreover, theoretical comparisons with the commonly used predictors are provided.

The remainder of the paper is organized as follows. In Section 2, we introduce the idea of Latin hypercube design-based block bootstrap and propose two methods to construct predictive distributions. In Section 3, the proposed predictors are shown to be unbiased. Theoretical comparisons with the regular bootstrap and the plugin approach are developed. In Section 4, finite-sample performance of the proposed methods is investigated by simulation studies. Applications to two real examples in computer experiments are given in Section 5. Discussions are given in Section 6.

1.2 GP prediction via design-based subsampling

1.2.1 Gaussian process models for computer experiments

Consider a computer experiment which has n inputs $\boldsymbol{x} \in R^d$ and produces output $y(\boldsymbol{x})$. To analyze the experiments, $y(\boldsymbol{x})$ is assumed to be a realization from a stochastic process

$$Y(\boldsymbol{x}) = \mu(\boldsymbol{x}) + Z(\boldsymbol{x}), \tag{1.1}$$

where the mean function is defined as $\mu(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}$ and $Z(\boldsymbol{x})$ is a stationary Gaussian process with mean 0 and covariance function $\sigma^2\psi$. The covariance function is defined as $cov\{Y(\boldsymbol{x}+\boldsymbol{h}), Y(\boldsymbol{x})\} = \sigma^2\psi(\boldsymbol{h};\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of correlation parameters for the correlation function $\psi(\boldsymbol{h};\boldsymbol{\theta})$ and $\psi(\boldsymbol{h};\boldsymbol{\theta})$ is a positive semidefinite function with $\psi(\boldsymbol{0};\boldsymbol{\theta}) = 1$ and $\psi(\boldsymbol{h};\boldsymbol{\theta}) = \psi(-\boldsymbol{h};\boldsymbol{\theta})$. Note that we assume the variables in the mean function are known and such a model is also known as universal kriging. However, the proposed framework is not limited to this assumption. It can be further extended to incorporate various variable selection methods for GP models (Chu et al. 2011, Li and Sudjianto 2005).

Suppose n realizations are observed and denoted by

$$\mathcal{D}_n = \{\big(\boldsymbol{x}_{t_1}, y(\boldsymbol{x}_{t_1})\big), \ldots, \big(\boldsymbol{x}_{t_n}, y(\boldsymbol{x}_{t_n})\big)\} = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}.$$

Let $\boldsymbol{y}_n = (y_1, \ldots, y_n)^T$, $\boldsymbol{X}_n = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$, $\boldsymbol{\phi} = (\theta^T, \boldsymbol{\beta}^T, \sigma^2)^T$ be the vector of all the parameters, and Θ be the parameter space. Based on (1.1), the likelihood function can be written as

$$f(\boldsymbol{y}_n, \boldsymbol{X}_n; \boldsymbol{\phi}) = \frac{|R_n(\boldsymbol{\theta})|^{-1/2}}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta})^T R_n^{-1}(\boldsymbol{\theta})(\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta})\},$$

where $R_n(\boldsymbol{\theta}) = [\psi(y(\boldsymbol{x}_i), y(\boldsymbol{x}_j); \boldsymbol{\theta}), i, j = 1, \ldots, n]$ is an $n \times n$ correlation matrix. Thus, the log-likehood function, ignoring a constant, is

$$\ell(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi}) = -\frac{1}{2\sigma^2}(\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta})^T R_n^{-1}(\boldsymbol{\theta})(\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}) - \frac{1}{2}\log|R_n(\boldsymbol{\theta})| - \frac{n}{2}\log(\sigma^2),$$

Here, the parameters $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and σ are unknown. They are estimated by likelihood-based methods such as maximum likelihood or restricted maximum likelihood (REML) (Irvine et al. 2007). In this paper, we focus on the study of maximum likelihood estimators and the results can be further extended to REML.

For a GP model, the maximum likelihood estimators (MLEs) can be obtained by

$$\hat{\boldsymbol{\beta}}_n = (\boldsymbol{X}_n^T R_n^{-1}(\boldsymbol{\theta})\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n^T R_n^{-1}(\boldsymbol{\theta})\boldsymbol{y}_n, \tag{1.2}$$

$$\hat{\sigma}_n^2 = (\boldsymbol{y}_n - \boldsymbol{X}_n\hat{\boldsymbol{\beta}}_n)^T R_n^{-1}(\boldsymbol{\theta})(\boldsymbol{y}_n - \boldsymbol{X}_n\hat{\boldsymbol{\beta}}_n)/n, \tag{1.3}$$

and

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}}\{n\log(\hat{\sigma}_n^2) + \log|R_n(\boldsymbol{\theta})|\}, \tag{1.4}$$

where $|R_n(\boldsymbol{\theta})|$ is the determinant of matrix $R_n(\boldsymbol{\theta})$.

Based on the MLEs, we are interested in predicting y_{n+1} at an untried new input $\boldsymbol{x}_{n+1}$ and quantifying the uncertainty. To achieve this, the conventional plug-in method predicts y_{n+1} by a distribution $g(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{Y}_n, \hat{\phi}_n)$ which is normally distributed with mean

$$\mu(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n) = \boldsymbol{x}_{n+1}^T\hat{\boldsymbol{\beta}}_n + \gamma_n(\hat{\boldsymbol{\theta}}_n)^T R_n^{-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_n - \boldsymbol{X}_n\hat{\boldsymbol{\beta}}_n) \tag{1.5}$$

and variance

$$\sigma^2(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n) = \hat{\sigma}_n^2\{1 - \gamma_n(\hat{\boldsymbol{\theta}}_n)^T R_n^{-1}(\hat{\boldsymbol{\theta}}_n)\gamma_n(\hat{\boldsymbol{\theta}}_n)\}, \tag{1.6}$$

where $\gamma_n(\hat{\boldsymbol{\theta}}_n)$ is the correlation between the new observation and the existing data, i.e. $\gamma_n(\hat{\boldsymbol{\theta}}_n) = \big[\psi(\boldsymbol{x}_i - \boldsymbol{x}_{n+1}; \hat{\boldsymbol{\theta}}_n), i = 1, \ldots, n\big]$.

Such a predictor is often computationally infeasible for massive data because it requires manipulations of a $n \times n$ correlation matrix $R_n(\hat{\boldsymbol{\theta}}_n)$, such as the calculation of $R_n^{-1}(\boldsymbol{\theta})$ and $|R_n(\boldsymbol{\theta})|$, which is computationally intensive and often intractable due to numerical issues. It is particularly difficult for massive data (i.e., large n) collected on irregular grids because Kronecker product techniques cannot be utilized to simplify the computation (Rougier 2008). Alternatives, such as Bayesian methods suffer from the same difficulty. Furthermore, the resulting plug-in predictors tend to underestimate the uncertainty because the variance in (1.6) is obtained by substituting the true parameters by their estimators. (Castrilln-Cands et al., 2016)

1.2.2 Latin hypercube design-based block subsampling

The main idea to achieve efficient computational reduction in GP estimation and prediction is to incorporate a new version of bootstrap subsampling called Latin hypercube design (LHD)-based block subsampling. A similar idea is first introduced by Zhao et al. (2018) to address the computation issue in variable selection. Inspired by its computational efficiency, we developed a new framework to construct predictive distributions.

The idea of bootstrap subsampling is attractive in many applications to achieve computational reductions, but direct applications with random subsamples is not efficient in GP estimation and prediction because of two reasons. First, it is known in the experimental design literature that the estimation efficiency of simple random sampling can be improved by certain stratification such as Latin hypercube designs (Mckay et al. 1979). Second, it is shown by Zhu and Stein (2006) that including clusters of points are important for capturing the local behavior of the process especially when the parameters are unknown in GP.

The LHD-based block subsampling has the following advantages. First, because of the one-dimensional balance property inherited from LHDs, the subsamples can spread out uniformly over the complete data and therefore the resulting subsamples are more representative. Second, the estimation and prediction calculated by the LHD-based subsamples are expected to outperform those from simple random samples because of the well-developed understanding of variance reduction in LHD compared with simple random sampling (MaKay et al. 1979). Third, the clusters of points within the blocks captures the local behavior of the process and therefore improves the estimation accuracy for correlation parameters which is essential for GP prediction.

The LHD-based subsampling can be described in the following three steps.

Step 1: *Denote the* d*-dimensional input space by* $\Gamma \in [0, l]^d$*. Divide each dimension into* m *equally spaced intervals so that* Γ *consists of* m^d *disjoint hypercubes/blocks. Define each block by mapping* $\boldsymbol{i}$ *to a* d*-dimensional hypercube*

$$\mathcal{B}_n(\boldsymbol{i}) = \{\boldsymbol{x} \in R^d : bi_j \leq x_j \leq b(i_j + 1) \text{ and } j = 1, ..., d\},$$

where $\boldsymbol{i} = (i_1, ...i_d)$*,* $i_j \in (0, ..., m-1)$*, represents the index of each block and* $b = l/m$ *is the edge length of the hypercube. Let* $|\mathcal{B}_n(\boldsymbol{i})|$ *be the number of observations in the* $\boldsymbol{i}$*th block. For simplicity, assume* $|\mathcal{B}_n(\boldsymbol{i})| = n/m^d$ *and the data points are equally spaced.*

Step 2: *Select* m *hypercubes according to a randomly generated* m*-run LHD, in which each column of the design matrix is a random permutation of* $\{0, \ldots, m-1\}$*. Denote the design points by* d*-dimensional vectors* $\boldsymbol{i}_1^*, \ldots, \boldsymbol{i}_m^*$ *and the corresponding selected blocks are* $\mathcal{B}_n(\boldsymbol{i}_1^*), \ldots, \mathcal{B}_n(\boldsymbol{i}_m^*)$*. The bootstrapped subsamples, denoted by* $y_1^*(\boldsymbol{x}_1^*), \ldots, y_N^*(\boldsymbol{x}_N^*)$*, are the observations in the selected blocks, where* $N = \sum_{i=1}^m |\mathcal{B}_n(\boldsymbol{i}_i^*)|$*. Based on the subsamples, maximum likelihood estimators* $\hat{\boldsymbol{\phi}}_N^*$ *can be obtained by (1.2)-(1.4).*

Step 3: *Repeat the second step* U *times to obtain bootstrapped MLEs* $\hat{\boldsymbol{\phi}}_{N(1)}^*, \cdots, \hat{\boldsymbol{\phi}}_{N(U)}^*$*. Based on these estimators, the bootstrap predictive distributions can be constructed using the methods described in Section 1.2.3.*

To illustrate the subsampling idea, we look at a simple example of an 6-run 2-dimensional LHD on the left panel of Figure 1. The design points are denoted by $\boldsymbol{i}_1^* = (0, 4)$, $\boldsymbol{i}_2^* = (1, 0)$, $\boldsymbol{i}_3^* = (2, 2)$, $\boldsymbol{i}_4^* = (3, 5)$, $\boldsymbol{i}_5^* = (4, 1)$, and $\boldsymbol{i}_6^* = (5, 3)$. On the right panel, consider

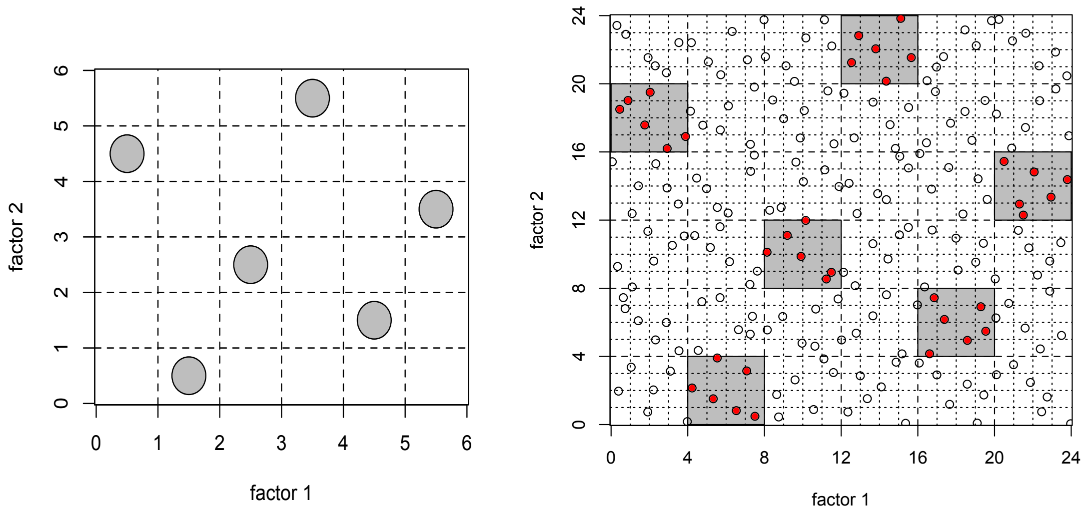


Figure 1.2.1: An example of LHD-based block bootstrap

$\Gamma \in [0, 24]^2$ with $d = 2$ and $l = 24$. The circles represent the settings in which computer experiments are performed and the total sample size is $n = 216$. According to the LHD on the left, we have $m = 6$, $b = 4$, $|\mathcal{B}_n(\boldsymbol{i})| = 6$. The corresponding LHD-based blocks are the six gray boxes on the right panel and the red dots are the resulting subsamples with size $N = 36$.

Note that, by the application of LHD-based block subsampling, the complexity is reduced from $O(n^3)$ to $O(n^3/m^{3(d-1)})$, which is particularly useful for high-dimensional problems in computer experiments. This method also allows parallel computing for large datasets. Furthermore, the assumption $|\mathcal{B}_n(\boldsymbol{i})| = n/m^d$ is mainly for notation simplicity and the procedure can be applied as long as the number of observations in each block is in the same order, i.e. $|\mathcal{B}_n(\boldsymbol{i}_i^*)| = O(n/m^d)$. We refer to Zhao et al. (2018) for various modifications to account for practical situations such as irregular design space (Draguljić et al. 2012, Hung et al. 2012), empty blocks, and the applications to a subset of variables.

1.2.3 Two construction methods for predictive distribution

To construct a predictive distribution based on the LHD-based subsamples, we developed two bootstrap procedures. One is called the *direct density prediction* method and the other is called the *Normal approximation* method. Both procedures utilize the LHD-based subsamples to construct predictive distributions; therefore, compared with using full data, the computational complexity is reduced. The difference between these two methods is how the Normal assumption is imposed. The direct density method imposes the Normal assumption on each bootstrap iteration which leads to the final predictive distribution following Normal mixture. On the other hand, the Normal approximation method assumes that the final predictive distribution is Normal and the mean and variance are estimated by LHD-based subsamples. The mathematical definition of the two methods are given as follows.

Definition 1 (Direct density prediction)**.** *Given the realization* $\{\boldsymbol{X}_n, \boldsymbol{y}_n\}$*, let* $\{\boldsymbol{X}_N^*, \boldsymbol{y}_N^*\}$ *be a bootstrap sample with empirical distribution* P^* *and* $\hat{\boldsymbol{\phi}}_N^*$ *be the maximiser of the log-likelihood* $\ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \boldsymbol{\phi})$*, a bootstrap predictive distribution is defined by*

$$g^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n) = \int g(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_N^*) dP^*(\boldsymbol{X}_N^*, \boldsymbol{y}_N^* \mid \boldsymbol{X}_n, \boldsymbol{y}_n), \tag{1.7}$$

where $g(\cdot)$ *is probability density function of Normal distribution with mean* $\mu(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)$ *and variance* $\sigma^2(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)$.

Based on the LHD-based subsamples, a Monte Carlo estimate of (1.7) can be obtained by

$$\tilde{g}^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \mathbf{y}_n) = U^{-1} \sum_{u=1}^{U} g(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_{N(u)}, \boldsymbol{Y}^*_{N(u)}, \hat{\phi}^*_{N(u)}),$$

where $\hat{\phi}^*_{N(u)}$ and $u = 1, ..., U$ are the MLEs obtained from each subsample. The resulting $\tilde{g}^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \mathbf{y}_n)$ follows a mixture distribution. When $U \to \infty$, $\tilde{g}^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \mathbf{y}_n)$ converges to $g^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n)$.

The conventional predictive distribution discussed in Section 1.2.1 is Normal, therefore a reasonable alternative is to assume a Normally distributed predictive distribution with mean and variance estimated as follows.

Definition 2 (Normal approximation)**.** *The predictive distribution is Normal with mean*

$$\mu^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n) = \int \mu(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}^*_N) dP^*(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N \mid \boldsymbol{X}_n, \boldsymbol{y}_n) \tag{1.8}$$

and variance

$$\sigma^{2*}(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n) = \int \sigma^2(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}^*_N) dP^*(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N \mid \boldsymbol{X}_n, \boldsymbol{y}_n). \tag{1.9}$$

Based on the LHD-based subsamples, the Monte Carlo estimates of (1.8) and (1.9) can be obtained by:

$$\tilde{\mu}^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n) = U^{-1} \sum_{u=1}^{U} \mu(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_{N(u)}, \boldsymbol{y}^*_{N(u)}, \hat{\phi}^*_{N(u)})$$

and

$$\tilde{\sigma}^{2*}(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n) = U^{-1} \sum_{u=1}^{U} \sigma^2(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_{N(u)}, \boldsymbol{y}^*_{N(u)}, \hat{\phi}^*_{N(u)}).$$

When $U \to \infty$, $\tilde{\mu}^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n)$ converges to $\mu^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n)$ and $\tilde{\sigma}^{2*}(\boldsymbol{x}_{n+1} \mid$

$\boldsymbol{X}_n, \boldsymbol{y}_n)$ converges to $\sigma^{2*}(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n)$.

1.3 Theoretical properties and comparisons

Theoretic properties including the unbiasedness and the variance of the proposed predictors are derived in this section. The results herein focus only on GP prediction, assuming that the estimator $\hat{\boldsymbol{\phi}}_N^*$ converges to the original MLE $\hat{\boldsymbol{\phi}}_n$ in probability which is shown by Zhao et al. (2018). Note that there are two distinct asymptotics, the fixed-domain (Stein 1999) and increasing domain (Cressie 1993, Mardia and Marshall 1984) asymptotics. However, theoretical results under fixed-domain asymptotics are limited in the literature due to its complex correlation structure in general (Ying 1993, Zhang 2004). It is shown by Zhang and Zimmerman (2005) that, given quite different behavior under the two frameworks in a general setting, their approximation quality performs about equally well for the exponential correlation function under certain assumptions. Therefore, we focus here on the increasing domain asymptotics as a fundamental step to provide insights about the bootstrap estimators.

We first construct an asymptotic expansion of the predictive distributions, which is a fundamental tool for the theoretical developments of the proposed method. Define the information matrix of the bootstrapped likelihood function evaluated at $\hat{\boldsymbol{\phi}}_n$ by

$$I = E^*\{-\nabla^2_{\boldsymbol{\phi}}\ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)\}$$

and I^{si} is the entry in the s^{th} row and i^{th} column of I^{-1}. The third-order derivative of the likelihood function evaluated at $\hat{\boldsymbol{\phi}}_n$ is then defined by

$$K_{ijk} = \frac{1}{2}E^*\{\frac{\partial^3\ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)}{\partial\phi_i\partial\phi_j\partial\phi_k}\}.$$

The cross products between the first and second order derivative of the predictive function and the second and third order derivative of the likelihood function evaluated at $\hat{\boldsymbol{\phi}}_n$ are

$$L^j_{s,i}(h) = E^*\{\frac{\partial h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)}{\partial \phi_s}\frac{\partial \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)}{\partial \phi_i}\frac{\partial \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)}{\partial \phi_j}\},$$

where $h_1(\boldsymbol{x}_{n+1} \mid ...) = I^{-1}h(\boldsymbol{x}_{n+1} \mid ...)$ and

$$J_{rs,ij}(h) = E^*\{\frac{\partial^2 h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)}{\partial \phi_r \phi_s}\frac{\partial \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)}{\partial \phi_i}\frac{\partial \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)}{\partial \phi_j}\},$$

and

$$M_{s,j,ik}(h) = \frac{1}{2}E^*\{\frac{\partial h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)}{\partial \phi_s}\frac{\partial \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)}{\partial \phi_j}\frac{\partial^2 \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)}{\partial \phi_i \partial \phi_k}\}.$$

The following theorem provides a third-order asymptotic expansion of the proposed predictive function. To facilitate the presentation, we use Einstein's summation convention hereafter: if an index appears twice in any one term, once as an upper and once as an lower index, summation over the index is applied.

Theorem 1. *Assume I is asymptotically nonsingular and the limit of $I^{-1/2}\nabla^2_{\boldsymbol{\phi}}\ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)I^{-1/2}$ is a unit matrix when $N \to \infty$. Then, the LHD-based bootstrap prediction function $h^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{Y}_n)$ has the following third-order asymptotic expansion:*

$$\begin{aligned} h^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{Y}_n) &= E^*h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n) \\ &+ I^{si}I^{jk}M_{s,j,ik} + \frac{1}{2}I^{ij}K_{irs}L^j_{r,s}(h) + I^{rj}I^{si}J_{rs,ij}(h) + O_{p^*}(N^{-2}). \end{aligned}$$

Due to the correlation between $\boldsymbol{x}_{n+1}$ and $\boldsymbol{X}_n$, the first term $E^*h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)$

does not always equal to $h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\boldsymbol{\phi}}_n)$. Assuming data independence, an important special case of Theorem 1, which agrees with the result in Fushiki et.al (2005, Theorem 1), is the following.

Corollary 1. *If $\psi(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0$ if $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, the LHD-based bootstrap prediction function $h^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{Y}_n) = h^*(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)$ has the following third-order asymptotic expansion:*

$$h^*(\boldsymbol{x}_{n+1}) = h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n) + I^{si}I^{jk}M_{s,j,ik} + \frac{1}{2}I^{ij}K_{irs}L^j_{r,s} + I^{rj}I^{si}J_{rs,ij} + O_{p^*}(N^{-2}).$$

Based on the asymptotic expansion in Theorem 1, we show that the two new predictors are unbiased and their variances can be rewritten as in the next theorem. Denote the predictive mean and variance of the direct density method by $\mu_1^*(\cdot)$ and $\sigma_1^{2*}(\cdot)$. Similarly, denote $\mu_2^*(\cdot)$ and $\sigma_2^{2*}(\cdot)$ for the normal approximation method. Let $\sum_{\boldsymbol{i}}$ be the summation of all m^d blocks and $\sum_{\boldsymbol{\pi}}$ be the summation of independent permutation over $\{0, 1, \ldots, m-1\}$.

Theorem 2. *Under the regularity conditions given in the Appendix, we have:*

i. The proposed predictors, μ_1^* and μ_2^*, are unbiased, i.e.,

$$\boldsymbol{E}\{\mu(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n) - \mu_1^*\} = \boldsymbol{E}\{\mu(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n) - \mu_2^*\} \quad \to 0$$

ii. The predictive variances have the following relationship:

$$\begin{aligned}
&P(\sigma_1^{2*} \geq \sigma_2^{2*}) \to 1,\\
&\sigma_1^{2*} = \sigma_2^{2*} + \frac{1}{(m!)^{d-1}}\sum_{\boldsymbol{\pi}}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 + o_p(1),\\
&\sigma_2^{2*} = \hat{\sigma}_n^2\Big\{1 - \frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)\Big\} + o_p(1).
\end{aligned}$$

The next theorem focuses on the comparison of the predictive variance of the plug-in predictive distribution defined in (1.6) and those of the two new predictors.

Theorem 3. *Under the regularity assumptions given in Appendix, we have*

$$P(\sigma_1^{2*} \geq \sigma^2(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n) \to 1,$$
$$P(\sigma_2^{1*} \geq \sigma^2(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n) \to 1.$$

It is known that the regular plug-in predictor interpolates the observed data. The next theorem shows that although this interpolation property cannot be guaranteed by the proposed predictors, the predictive variance on an existing data point is smaller than the variance on an untried point. For the direct density approach, denote the variance within the sampled data by $\sigma_1^{2*(I)}$ and the variance for out-of-sample by $\sigma_1^{2*(O)}$. Similarly, we have $\sigma_2^{2*(I)}$ and $\sigma_2^{2*(O)}$ for the normal approximation method.

Theorem 4. *Under the regularity assumptions given in Appendix, we have:*

(i). The in-sample predictive variances are

$$\begin{aligned}
\sigma_1^{2*(I)} &= \sigma_2^{2*(I)} + o_p(1) \\
&+ (1 - \frac{1}{m^{d-1}})\frac{1}{(m!)^{d-1}} \sum_{\boldsymbol{\pi}} [\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2, \\
\sigma_2^{2*(I)} &= (1 - \frac{1}{m^{d-1}})\hat{\sigma}_n^2[1 - \frac{1}{m^{d-1}} \sum_{\boldsymbol{i}} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)] + o_p(1).
\end{aligned}$$

(ii). Comparison of the in-sample and out-of-sample predictive variance:

$$\begin{aligned}\sigma_1^{2*(O)} - \sigma_1^{2*(I)} &= \sigma_2^{2*(O)} - \sigma_2^{2*(I)} + o_p(1) \\ &+ (mm!)^{1-d}\sum_{\boldsymbol{\pi}}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2, \\ \sigma_2^{2*(O)} - \sigma_2^{2*(I)} &= \frac{\hat{\sigma}_n^2}{m^{d-1}}[1 - \frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)] + o_p(1)\end{aligned}$$

i.e.

$$P(\sigma_1^{2*(O)} \geq \sigma_1^{2*(I)}) \to 1, \quad P(\sigma_2^{2*(O)} \geq \sigma_2^{2*(I)}) \to 1.$$

For Theorem 4, although the proposed predictors do not have the interpolation property, their in-sample predictive variances are in general smaller than their out-of-sample variances.

1.4 Simulation studies

The objective of this section is to demonstrate the finite-sample performance of the proposed methods. They are compared with some existing methods, including the regular GP model and the conventional bootstrap prediction. All the simulations are conducted by a 2.4 GHz Intel Core i5, 8GB 1600 MHz DDR3 workstation under Python 3.5.2 in MAC OS X.

1.4.1 Comparisons with regular MLE

The finite sample performance of the proposed methods are compared with the regular MLE using full data, denoted by "ALLData". Three different settings of LHD-based block bootstrap, $m = 3, 4$ and 5, are performed. The outputs are simulated from a Gaussian process with the mean function coefficients $\boldsymbol{\beta} = (2, -2, 1)$ and the correlation function

$$\psi(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp(-\sum_{i=1}^{3} |x_{1i} - x_{2i}|/\theta_i),$$

where $\theta_1 = \theta_2 = \theta_3 = 0.4$ and $\sigma = 1$. Two sample sizes, $n = 2000$ and 4000, are considered and the design points are generated from regular grid over the region $[0, 1]^3$. For each sample size, 50 training samples and 100 testing samples are generated. The performance of parameter estimation is summarized in Table 1.4.1 based on 100 replicates with 10 LHD-based block bootstrap samples implemented for the proposed method. In addition, the mean squared prediction errors (MSPE) for the testing datasets and the averaged computing time are both reported.

The results demonstrate that the estimated parameters using LHD-based block bootstrap are in general consistent with those obtained using the complete data. When $n = 2000$, the standard deviations increase with the number of blocks m, especially for the correlation parameters. This is not surprising because the sample sizes are smaller for larger m and ”AllData” implies the special case of ”m=1”. The impact of m on the estimation variance appears to be smaller when sample size increases to $n = 4000$. In terms of computing time, LHD-based block bootstrap is much faster than the conventional GP modeling especially for large n.

Note that the proposed methods are particularly useful for data collected from irregular grids. The reason to generate the simulations from a regular grid is that the MLE calculation using full data, under this setting, can be further simplified by Kronecker product techniques and some matrix singularity can be avoided (Rougier 2008). But these techniques are not applicable to datasets collected from an irregular grid. Therefore, the computational advantage of the proposed methods are expected to be even more significant

Table 1.4.1: Comparisons with regular MLE (standard deviation in parenthesis).

	AllData	LHD		
		m=3	m=4	m=5
		$n = 2000$		
θ_1	0.40(0.03)	0.43(0.09)	0.48(0.27)	0.90(2.68)
θ_2	0.40(0.03)	0.42(0.10)	0.45(0.24)	0.46(0.26)
θ_3	0.39(0.03)	0.45(0.13)	0.42(0.15)	0.50(0.41)
β_1	2.02(0.52)	2.04(0.68)	2.13(0.67)	2.06(0.72)
β_2	-2.04(0.57)	-1.98(0.70)	-2.03(0.64)	-2.00(0.73)
β_3	1.05(0.55)	1.03(0.69)	1.02(0.72)	1.04(0.68)
$MSPE$	0.10(0.14)	0.24(0.32)	0.33(0.46)	0.44(0.61)
$Time$	76.78(5.12)	10.96(4.03)	7.84(3.60)	4.83(1.67)
		$n = 4000$		
θ_1	0.40(0.02)	0.44(0.09)	0.43(0.13)	0.41(0.13)
θ_2	0.40(0.03)	0.44(0.09)	0.46(0.11)	0.41(0.14)
θ_3	0.40(0.02)	0.42(0.08)	0.44(0.12)	0.40(0.12)
β_1	2.07(0.53)	2.11(0.60)	2.10(0.68)	2.24(0.64)
β_2	-2.01(0.52)	-2.05(0.56)	-2.04(0.60)	-2.15(0.65)
β_3	1.04(0.49)	1.02(0.71)	1.02(0.60)	1.00(0.67)
$MSPE$	0.07(0.09)	0.16(0.22)	0.22(0.31)	0.27(0.38)
$Time$	605.83(35.93)	58.38(5.21)	20.53(4.41)	12.39(3.61)

for data collected from irrregular grids.

1.4.2 Comparisons of prediction uncertainty

We also compare the proposed predictive distributions with existing methods by looking at their predictive variance. Two existing methods, the regular bootstrap and the plugin predictive distribution, are considered. The regular bootstrap, although computationally expensive, can serve as a benchmark for capturing the true prediction uncertainty. Simulations are generated from the same model given in Section 1.4.1. Due to the computational constraints in regular bootstrap, we conduct relatively smaller sample sizes, $n = 1000$ and $n = 2000$, for comparison. The predictive variance is evaluated based on 100 untried settings with 50 replications. Both LHD-based subsampling and the regular bootstrap are performed using the two construction approaches, direct density and Normal approximation, and they are all calculated based on 10 resampling.

The performance on predictive variance is summarized in Table 1.4.2. The LHD-based subsampling is denoted by "LHD". In general, using LHD subsampling, the predictive variance constructed by direct density is larger than the one by Normal approximation, which is consistent with the theoretical results in Theorem 2. It is also not surprising to see that the predictive variances obtained from LHD subsampling are larger than the benchmark results from regular bootstrap, and the differences become smaller when sample size increases. Comparing with the plugin approach, the proposed methods provide a better quantification of the predictive uncertainty as expected. Furthermore, it is interesting to observe a significant amount of computational reduction even compared with the plug-in approach. So this result suggests that, by the LHD-based subsampling, the computational efforts can be effectively allocated to capture the uncertainty instead of large matrix manipulations.

Table 1.4.2: Comparisons of predictive variance (standard deviation in parenthesis).

	LHD (m=3)	Regular Bootstrap	Plugin
	$n = 1000$		
Direct Density	0.53(0.19)	0.35(0.15)	0.15(0.04)
Normal	0.41(0.08)	0.22(0.05)	
Time	5.50(2.09)	405.30(119.61)	19.11(6.45)
	$n = 2000$		
Direct Density	0.39(0.22)	0.33(0.16)	0.10(0.02)
Normal	0.31(0.08)	0.20(0.06)	
Time	10.96(4.03)	1917.15(543.02)	76.78(5.12)

1.5 Real Examples

1.5.1 A data center thermal management example

A data center is a computing infrastructure facility that houses large amounts of information technology equipment used to process, store, and transmit digital information. Data center facilities constantly generate large amounts of heat to the room, which must be maintained at an acceptable temperature for reliable operation of the equipment. A significant fraction of the total power consumption in a data center is for heat removal; therefore, determining the most efficient cooling mechanism has become a major challenge. To solve the problem, a crucial step is to model the thermal distribution at different experimental settings (Hung et al. 2012).

For a data center thermal study, physical experiments are not always feasible because some settings are highly dangerous and expensive to perform. Therefore, simulations based on computational fluid dynamics (CFD) are widely used. In this example, CFD simulations are conducted at IBM T. J. Watson Research Center based on a real data center layout. Detailed discussions about the CFD simulations can be found in (Lopez and

Hamann 2011). The first three columns in Table 1.5.1 list nine variables and their levels in the CFD simulations, including four computer room air conditioning (CRAC) units with different flow rates ($x_1, ..., x_4$), the overall room temperature setting (x_5), the perforated floor tiles with different percentage of open areas (x_6), and spatial location in the data center (x_7 to x_9). There are 27,000 temperatures simulated from the CFD simulator and these temperature outputs are obtained from an irregular grid over the 9-dimensional experimental space.

It is computationally intensive to build a GP model based on the complete CFD data. So we implement the proposed LHD-based block bootstrap approach with $m = 3$ for variables x_6, x_7 and x_9, which are the top three factors with highest levels. The fitted GP model is summarized by the last two columns of Table 1.5.1, where $\hat{\beta}$ represents the estimated mean function coefficients and $\hat{\theta}$ represents the correlation parameters estimated based on exponential covariance function. From the fitted model, it appears that the height (x_9) in a data center has a relatively larger effect, particularly to the mean function. Furthermore, we find the temperatures increase dramatically with height based on the predicted heat map at three different heights (Figure 1.5.1) with an untried setting (i.e., CRAC unit 1 flow rate 6500, unit 2 flow rate 6500, unit 3 flow rate 2750, unit 4 flow rate 2750, room temperature 70 (F) and tile percentage 59). These findings can be validated by the general understanding of thermal dynamics.

1.5.2 Ice sheet thickness modeling

The second application focuses on the study of ice sheet thickness using the community ice sheet model (CISM; Rutt et al. 2009, Price et al. 2011). The main objective of this model is to understand ice sheet behavior and its impact on climate. CISM mimics the effects of past climate on the current ice sheet state by considering a model of an idealized

Table 1.5.1: LHD Bootstrap analysis of thermal management data (standard deviation in parenthesis).

	Variable	Levels	$\hat{\beta}$	$\hat{\theta}$
x_1	CRAC unit 1 flow rate (cfm)	(0,7000,8500,10000 11500,13000)	-8.58(0.96)	0.85(0.17)
x_2	CRAC unit 2 flow rate (cfm)	(0,7000,8500,10000 11500,13000)	-11.12(1.26)	0.77(0.23)
x_3	CRAC unit 3 flow rate (cfm)	(0,2500,4000,5500)	-6.83(0.80)	1.14(0.27)
x_4	CRAC unit 4 flow rate (cfm)	(0,2500,4000,5500)	-6.26(0.98)	1.70(0.71)
x_5	Room temperature setting (F)	(65,67,69,71,73, 75)	-0.82(0.66)	3.39(0.94)
x_6	Tile open area percentage (%)	$(15, 25, 35, 45$ $(55, 65, 75)$	0.15(3.63)	1.24(0.91)
x_7	Location in x-axis	8 unequally spaced	-5.09(2.72)	0.14(0.11)
x_8	Location in y-axis	4 unequally spaced	3.70(2.18)	0.62(0.25)
x_9	Height	18 equally spaced	33.43(3.90)	21.61(0.22)

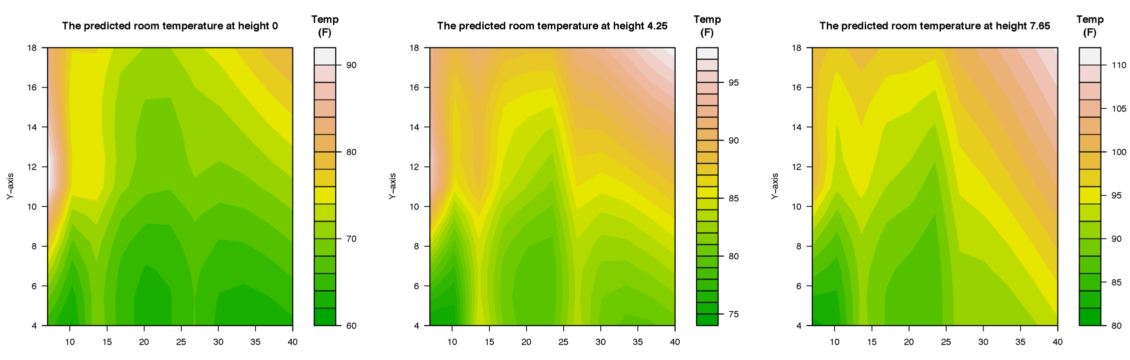


Figure 1.5.1: Bootstrap predictive heat map in a data center

ice sheet over a rectangular region which is flowing out to sea on one side, while accumulating ice from prescribed precipitation over a time of 1000 years. There are two control variables in CISM, one is a constant term in the Glen-Nye flow law (Greve and Blatter 2009) controlling the deformation of the ice sheet denoted by x_1 and the other controls the heat conductivity in the ice sheet denoted by x_2. The simulated thickness is produced on a 27×32 rectangular lattice of the spatial locations, denoted by x_3 and x_4. We focus on the central part of the icebergs by taking the middle 13×16 rectangular lattice in this analysis. A set of simulations with 20 different combinations of the two control variables is considered and therefore the total sample size is $n = 4160$. The detailed variable settings can be found in Higdon et al. (2013).

The focus of this study is to compare the performance of the proposed method with conventional GP using full data in real applications. A four-dimensional GP is considered for the analysis of simulation results from CISM. Estimation and prediction performance is evaluated based on a 10-fold cross validation. The LHD-based subsamples are obtained by the setting of ($m = 3$, $U = 10$) and each LHD-based subsample has size $N = 139$. The results are summarized in Table 1.5.2 with the conventional GP denoted by "Alldata" and the proposed method denoted by "LHD". RMSPE is the root mean squared prediction error calculated from the 10-fold cross validation.

From the results in Table 1.5.2, it appears that even with only $3.7\%(\approx 1/27)$ of the data in each subsample, the LHD-based approach can provide a reasonable performance in terms of parameter estimation and prediction. The computational time is reduced for more than 99.3% using the proposed method. In general, the estimation for x_3 seems to be more challenging than the other variables due to its relatively smaller effect. One example of the iceberg thickness prediction is demonstrated in Figure 1.5.2 over the entire spatial location with the parameter setting of $x_1 = 2.40$ and x_2=6.53×10^4. The left panel is the

Table 1.5.2: LHD Bootstrap analysis of CISM data (standard deviation in parenthesis).

	Alldata		LHD	
	$\hat{\beta}$	$\hat{\theta}$	$\hat{\beta}$	$\hat{\theta}$
x_1	-0.80(4.0×10^{-3})	7.26(11.00)	-0.84(0.08)	23.71(19.56)
x_2	0.31(6.0×10^{-3})	3.41(8.15)	0.42(0.05)	3.82(4.00)
x_3	7.3×10^{-5}(6.0×10^{-4})	0.03(9.5×10^{-3})	-1.2 $\times 10^{-4}$(2.2×10^{-3})	3.4×10^{-3}(1.1×10^{-3})
x_4	0.11(4.6×10^{-4})	0.65(0.21)	0.08(0.01)	0.21(0.58)
RMSPE	0.01		0.07	
Time (Sec)	10097.39		66.36	

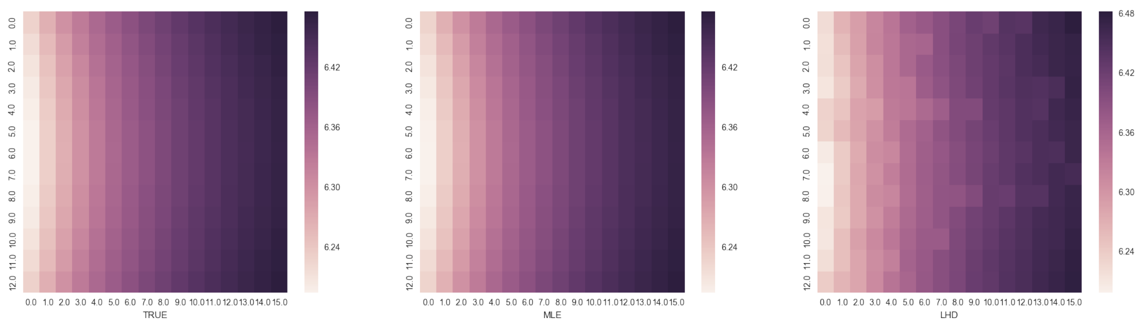


Figure 1.5.2: Thickness predictions of icesheet. Left: Truth. Middle: Prediction using conventional GP. Right: LHD-based method.

original simulation outputs from CISM. The middle panel is the plug-in prediction using full data. The right panel is the prediction obtained from the LHD-based approach. It shows that, given some roughness due to the small subsample size, the prediction using LHD-based approach can efficiently capture the underlying structure.

1.6 Discussion

We present a LHD-based block subsampling procedure with two prediction methods to tackle the computational difficulties and uncertainty quantification issues in GP prediction. The new procedure borrows the strength of space-filling designs to provide an efficient subsampling plan and reduce computational complexity. Theoretical properties of the proposed predictive distributions are discussed. The proposed procedure is applied to two complex computer experiments with high dimensional inputs and massive outputs.

Future work will be in the following directions. First, extensions of the proposed procedure to optimal designs with better space-filling properties are intuitively appealing. For example, it is known that randomly generated LHDs can contain some structure. To further enhance desirable space-filling properties, various modifications are proposed. Numerical comparisons and theoretical developments of the generalization to different types of optimal space-filling designs will be carefully studied. Second, an interesting and important issue of the LHD-based block bootstrap is to determine the optimal block size. This topic has been discussed for conventional block bootstrap methods (Hall et al. 1995, Lahiri 1999, Nordman et al. 2007), however the solutions therein are not directly applicable to GP models. We plan to study the optimal block size for the propose procedure based on some new criteria defined for GP.

1.7 Technical proofs

1.7.1 Assumptions

Let $\boldsymbol{y_i} = (y_s(\boldsymbol{x}_s), \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}))$ and $\boldsymbol{X_i} = (\boldsymbol{x}_s, \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}))^T$ denote the data in the ith block. Define $R_{\boldsymbol{i},\boldsymbol{j}}(\boldsymbol{\theta})$=$\left[\psi(y(\boldsymbol{x}_s), y(\boldsymbol{x}_t); \boldsymbol{\theta}),\ \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}), \boldsymbol{x}_t \in \mathcal{B}_n(\boldsymbol{j})\right]$, $D_n(\boldsymbol{\theta}) = \mathrm{diag}(R_{\boldsymbol{i},\boldsymbol{i}}(\boldsymbol{\theta}))$ with $\boldsymbol{i} = (i_1, \ldots, i_d)$ in lexicographical order, and $E_n(\boldsymbol{\theta}) = R_n(\boldsymbol{\theta}) - D_n(\boldsymbol{\theta})$. We need the following assumptions for the proposed prediction procedures.

1. $m = o(n^{1/d})$ and $m \to \infty$.

2. $\lim_{n\to\infty} \sup_{\boldsymbol{\theta}} \lambda_{\max}(E_n(\boldsymbol{\theta})) = 0$, when the block space $b = l/m \to \infty$.

These two assumptions are also the necessary conditions for the consistency of the bootstrap estimators (Zhao et al. 2018). The second assumption aims to control the correlation between bootstrap blocks.

1.7.2 Proof of Theorem 1

By definition, the bootstrap predictive function is

$$\begin{aligned} h^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n) = &\ E^*\{h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_N^*)\} \\ = &\ \int h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_N^*) dP^*(\boldsymbol{X}_N^*, \boldsymbol{y}_N^* \mid \boldsymbol{X}_n, \boldsymbol{y}_n). \end{aligned}$$

Take Taylor expansion of $h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_N^*)$ at $\hat{\boldsymbol{\phi}}_n$ for each bootstrap, we have

$$\begin{aligned} & h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_N^*) \\ = & h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) + \nabla_{\boldsymbol{\phi}}^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n)(\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n) \\ & + \frac{1}{2}(\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n)^T \nabla_{\boldsymbol{\phi}}^2 h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n)(\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n) \\ & + \frac{1}{6}\nabla\{(\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n)^T \nabla_{\boldsymbol{\phi}}^2 h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n)(\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n)\}(\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n) \\ & + O_{p^*}(\|\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n\|_2^4). \end{aligned}$$

So we only need to calculate the expectation of each term on the right-hand side of the equation above.

Again, we treat $\nabla\ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_N^*)$ as a function of $\boldsymbol{\phi}$ and take the second order Taylor expansion at $\hat{\boldsymbol{\phi}}_n$. Recall that $\nabla\ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_N^*) = 0$ and

$$0 = \nabla\ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) + \nabla^2\ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n)\boldsymbol{\omega} + \frac{1}{2}\nabla\{\nabla^2\ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n)\boldsymbol{\omega}\}\boldsymbol{\omega} + O_{P^*}(\|\boldsymbol{\omega}\|_2^3) \tag{1.10}$$

where $\boldsymbol{\omega} = \hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n$. Multiplying (1.10) by $\nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n)$, we have

$$
\begin{aligned}
0 =& \nabla \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \\
&+ \nabla^2 \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \\
&+ \frac{1}{2} \nabla \{\nabla^2 \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \boldsymbol{\omega}\} \boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) + O_{P^*}(\|\boldsymbol{\omega}\|_2^3). \quad (1.11)
\end{aligned}
$$

Using the fact that $\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n = I^{-1} \nabla \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) + O_{P^*}(N^{-1/2})$, take expectations of each term in the equation above. For presentation simplicity, $\boldsymbol{X}_N^*, \boldsymbol{y}_N^*$ are omitted in the calculation below.

$$
\begin{aligned}
& E^*\{\nabla^2 \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n)\} \\
= \quad & E^*\{\nabla^2 \ell(\hat{\boldsymbol{\phi}}_n)\} E^*\{\boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} + Cov^*\{\nabla^2 \ell(\hat{\boldsymbol{\phi}}_n), \boldsymbol{\omega} \nabla^T \nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \\
= \quad & -I E^*\{\boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} + I^{-1} E^*\{\nabla^2 \ell(\hat{\boldsymbol{\phi}}_n) \nabla \ell(\hat{\boldsymbol{\phi}}_n) \nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \\
& -I^{-1} E^*\{\nabla^2 \ell(\hat{\boldsymbol{\phi}}_n)\} E^*\{\nabla \ell(\hat{\boldsymbol{\phi}}_n) \nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\}
\end{aligned}
$$

Using the same technique, we have

$$
\begin{aligned}
& E^*\{\nabla\{\nabla^2 \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \boldsymbol{\omega}\} \boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n)\} \\
= \quad & (K_{irs} L_{r,s}^j)_{i,j=1,\ldots,N} + O_{P^*}(N^{-2}).
\end{aligned}
$$

Plugging the equations back into (1.11), we have

$$\begin{aligned}0 =&E^*\{\nabla\ell(\hat{\boldsymbol{\phi}}_n)\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} - IE^*\{\boldsymbol{\omega}\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \\ &+ I^{-1}E^*\{\nabla^2\ell(\hat{\boldsymbol{\phi}}_n)\nabla\ell(\hat{\boldsymbol{\phi}}_n)\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \\ &- I^{-1}E^*\{\nabla^2\ell(\hat{\boldsymbol{\phi}}_n)\}E^*\{\nabla\ell(\hat{\boldsymbol{\phi}}_n)\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \\ &+ \frac{1}{2}(K_{irs}L^j_{r,s})_{i,j=1,\ldots,N} + O_{P^*}(N^{-2}).\end{aligned}$$

Thus,

$$\begin{aligned}E^*\{\boldsymbol{\omega}\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \;=\;& I^{-2}E^*\{\nabla^2\ell(\hat{\boldsymbol{\phi}}_n)\nabla\ell(\hat{\boldsymbol{\phi}}_n)\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \\ &+\frac{1}{2}I^{-1}(K_{irs}L^j_{r,s})_{i,j=1,\ldots,N} + O_{P^*}(N^{-2}).\end{aligned}$$

Taking trace on both side of the equation, we have

$$E^*\{\boldsymbol{\omega}\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} = I^{si}I^{jk}M_{s,j,ik} + \frac{1}{2}I^{ij}I^{jk}K_{irs}L^j_{r,s}(h) + O_{P^*}(N^{-2}).$$

Similarly,

$$E^*(\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n)^T\nabla^2_{\boldsymbol{\phi}} h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)(\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n) = \{I^{rj}I^{si}J_{rs,ij}(h)\} + O_{P^*}(N^{-2}).$$

The result follows by plugging the two equations into the Taylor expansion of $h^*(\cdot)$. $\square$

1.7.3 Proof of Theorem 2

To investigate the asymptotic properties of the predictions from LHD-based block bootstrap, we decompose the likelihood function into blocks. For each block, denote $\boldsymbol{y}_{\boldsymbol{i}} = (y_s(\boldsymbol{x}_s), \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}))$, $\boldsymbol{X}_{\boldsymbol{i}} = (\boldsymbol{x}_s, \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}))^T$, $R_{\boldsymbol{i},\boldsymbol{j}}(\boldsymbol{\theta})$=$\big[\psi(y(\boldsymbol{x}_s), y(\boldsymbol{x}_t); \boldsymbol{\theta}),$

$\boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}), \boldsymbol{x}_t \in \mathcal{B}_n(\boldsymbol{j})]$ and $\boldsymbol{z}_{\boldsymbol{i}} = R_{\boldsymbol{i},\boldsymbol{i}}^{-1/2}(\boldsymbol{\theta})(\boldsymbol{y}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{i}}\boldsymbol{\beta})$. Then, we can rewrite the normalised log-likelihood function $n^{-1}\ell(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi})$ as

$$\begin{aligned} Q_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi}) = & -(2n\sigma^2)^{-1} \textstyle\sum_{s=1}^n z_s^2 - (2n)^{-1} \sum_{s=1}^n \log(\lambda_s) \\ & -(2n)^{-1} \textstyle\sum_{s=1}^n \log(\sigma^2) + n^{-1} r_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi}) \\ = & \; n^{-1} \textstyle\sum_{s=1}^n q_s(\omega, \boldsymbol{\phi}) + n^{-1} r_n(\omega, \boldsymbol{\phi}), \end{aligned}$$

where $\{\lambda_s, s = 1, \ldots, n\} = \{\text{eigenvalues of } |R_{\boldsymbol{i},\boldsymbol{i}}(\boldsymbol{\theta})|, \boldsymbol{i} = (i_1, \ldots, i_d)\}$ with $(i_1, \ldots, i_d)$ in lexicographical order and eigenvalues from the largest to the smallest. Note that $r_n(\omega, \boldsymbol{\phi})$ $= \ell(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi}) - \sum_{s=1}^n q_s(z_s, \boldsymbol{\phi})$ contains all terms involving the off block-diagonal terms. Define $D_n(\boldsymbol{\theta}) = \text{diag}(R_{\boldsymbol{i},\boldsymbol{i}}(\boldsymbol{\theta}))$ and $E_n(\boldsymbol{\theta}) = R_n(\boldsymbol{\theta}) - D_n(\boldsymbol{\theta})$. Assuming that $E_n(\boldsymbol{\theta}) = U_n(\boldsymbol{\theta})U_n^T(\boldsymbol{\theta})$, we have

$$\begin{aligned} r_n(\omega, \boldsymbol{\phi}) \;=\; & \frac{1}{2\sigma^2(1+g)}(\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta})^T D_n^{-1}(\boldsymbol{\theta}) E_n(\boldsymbol{\theta}) D_n^{-1}(\boldsymbol{\theta})(\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}) \\ & + \frac{1}{2}\log|I_n + U_n^T(\boldsymbol{\theta}) D_n^{-1}(\boldsymbol{\theta}) U_n(\boldsymbol{\theta})|, \end{aligned}$$

where $g = \text{trace}(E_n(\boldsymbol{\theta})D_n^{-1}(\boldsymbol{\theta}))$.

The maximum likelihood estimator is obtained by $\hat{\boldsymbol{\phi}}_n = \arg\max_{\boldsymbol{\phi}} Q_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi})$. Analogue to the decomposition for $Q_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi})$, the log-likelihood function for LHD-based block bootstrap samples can be written as

$$Q_N^*(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \boldsymbol{\phi}) \;=\; N^{-1} \sum_{s=1}^N q_s^*(\cdot, \omega, \boldsymbol{\phi}) + N^{-1} r_N^*(\cdot, \omega, \boldsymbol{\phi}), \tag{1.12}$$

where $r_N^*(\cdot, \omega, \boldsymbol{\phi})$ contains all terms involving the off block-diagonal terms with bootstrapped samples. Specifically,

$$\begin{aligned} r_N^*(\cdot, \omega, \boldsymbol{\phi}) &= \frac{1}{2\sigma^2(1+g^*)}(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\boldsymbol{\beta})^T {D_N^*}^{-1}(\boldsymbol{\theta}) E_N^*(\boldsymbol{\theta}) {D_N^*}^{-1}(\boldsymbol{\theta})(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\boldsymbol{\beta}) \\ &\quad + \frac{1}{2}\log|I_N + {U_N^*}^T(\boldsymbol{\theta}) {D_N^*}^{-1}(\boldsymbol{\theta}) U_N^*(\boldsymbol{\theta})|, \end{aligned}$$

where $D_N^*(\boldsymbol{\theta}) = \mathrm{diag}(R_{\boldsymbol{i}_j^*, \boldsymbol{i}_j^*}(\boldsymbol{\theta}), j = 1, \ldots, m)$ and $E_N^*(\boldsymbol{\theta}) = R_N^*(\boldsymbol{\theta}) - D_N^*(\boldsymbol{\theta})$ with $E_N^*(\boldsymbol{\theta}) = U_N^*(\boldsymbol{\theta}) {U_N^*}^T(\boldsymbol{\theta})$; $g^* = \mathrm{trace}(E_N^*(\boldsymbol{\theta}) {D_N^*}^{-1}(\boldsymbol{\theta}))$. The bootstrapped version of $\hat{\boldsymbol{\phi}}_n$ is $\hat{\boldsymbol{\phi}}_N^* = \arg\max_{\boldsymbol{\phi}} Q_N^*(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \boldsymbol{\phi})$, which is a consistent estimate of $\hat{\boldsymbol{\phi}}_n$ according to Zhao et al. (2018).

Similar to the decomposition of the bootstrapped likelihood (1.12), we rewrite the weighted average of the bootstrapped data. Recall $D_N^*(\hat{\boldsymbol{\theta}}_n) = \mathrm{diag}(R_{\boldsymbol{i}_j^*, \boldsymbol{i}_j^*}(\hat{\boldsymbol{\theta}}_n), j = 1, \ldots, m)$ and $E_N^*(\hat{\boldsymbol{\theta}}_n) = R_N^*(\hat{\boldsymbol{\theta}}_n) - D_N^*(\hat{\boldsymbol{\theta}}_n)$ with $E_N^*(\hat{\boldsymbol{\theta}}_n) = U_N^*(\hat{\boldsymbol{\theta}}_n) {U_N^*}^T(\hat{\boldsymbol{\theta}}_n)$; $\hat{g}^* = \mathrm{trace}(E_N^*(\hat{\boldsymbol{\theta}}_n) {D_N^*}^{-1}(\hat{\boldsymbol{\theta}}_n))$. Then $\gamma_N^*(\hat{\boldsymbol{\theta}}_n)^T R^{*-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\hat{\boldsymbol{\beta}}_n)$ can be written as

$$\begin{aligned} & \gamma_N^*(\hat{\boldsymbol{\theta}}_n)^T R^{*-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\hat{\boldsymbol{\beta}}_n) \\ = & \sum_{j=1}^m \gamma_{\boldsymbol{i}_j^*}(\hat{\boldsymbol{\theta}}_n)^T R^{-1}_{\boldsymbol{i}_j^*, \boldsymbol{i}_j^*}(\hat{\boldsymbol{\theta}}_n)(y_{\boldsymbol{i}_j^*} - \boldsymbol{X}_{\boldsymbol{i}_j^*}\hat{\boldsymbol{\beta}}_n) + s_N^*(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n), \end{aligned}$$

where $\gamma_N^*(\hat{\boldsymbol{\theta}}_n)$ is the correlation between $\boldsymbol{x}_{n+1}$ and the bootstrapped data $\boldsymbol{X}_N^*$ calculated at $\hat{\boldsymbol{\theta}}_n$ and R^* is the correlation matrix of the bootstrapped data $\boldsymbol{X}_N^*$ calculated at $\hat{\boldsymbol{\theta}}_n$ as well; and $s_N^*(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n)$ contains all terms involving the off block-diagonal terms with bootstrapped samples. Specifically,

$$s_N^*(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n) = \frac{1}{(1+\hat{g}^*)}\gamma_N^*(\hat{\boldsymbol{\theta}}_n)^T {D_N^*}^{-1}(\hat{\boldsymbol{\theta}}_n) E_N^*(\hat{\boldsymbol{\theta}}_n) {D_N^*}^{-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\hat{\boldsymbol{\beta}}_n).$$

According to Theorem 1, for both direct density prediction method and normal prediction method, the predictive distribution has mean

$$
\begin{aligned}
& E^*\{\boldsymbol{x}_{n+1}^T\hat{\boldsymbol{\beta}}_n + \gamma_N^*(\hat{\boldsymbol{\theta}}_n)^T R^{*-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\hat{\boldsymbol{\beta}}_n)\} + o_p(1) \\
= \; & \boldsymbol{x}_{n+1}^T\hat{\boldsymbol{\beta}}_n + \frac{1}{m^{d-1}}\sum_{\boldsymbol{i}} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{i}}\hat{\boldsymbol{\beta}}_n) + E^*(s_N^*(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n)) + o_p(1).
\end{aligned}
$$

By the same treatment as the proof of $r_n(\cdot)$ and $r_N^*(\cdot)$ in Lemma 4 in Zhao et al. (2018), under condition A.3, we have $s_n(\cdot) = \frac{1}{(1+g)}\gamma_n(\hat{\boldsymbol{\theta}}_n)^T D_n^{-1}(\hat{\boldsymbol{\theta}})E_n(\hat{\boldsymbol{\theta}})D_n^{-1}\gamma_n(\hat{\boldsymbol{\theta}}_n) \to 0$ in P as well as $s_N^*(\cdot) \to 0$ prob-$P_{N,\omega}^*$ prob-P and $E^*(s_N^*(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n)) \to 0$ in P. Decompose the predictive mean of plug-in predictor using the same technique, we show that

$$
\begin{aligned}
& \boldsymbol{E}\{\mu(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \phi_n) - \mu_1^*\} = \boldsymbol{E}\{\mu(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \phi_n) - \hat{\mu}_2^*\} \\
= \; & \boldsymbol{E}\frac{m^{d-1}-1}{m^{d-1}}\sum_{\boldsymbol{i}} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{i}}\hat{\boldsymbol{\beta}}_n) + o_p(1) \\
\to \; & 0.
\end{aligned}
$$

where $\sum_{\boldsymbol{i}}$ is the summation of all m^d blocks.

The predictive distribution of direct density prediction method, which follows normal mixture, has variance

$$\begin{aligned}
\sigma_1^{2*} &= E^*\{\sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^{*(u)}, \boldsymbol{y}_N^{*(u)}, \hat{\phi}_n) + [\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^{*(u)}, \boldsymbol{y}_N^{*(u)}, \hat{\phi}_n) \\
&\quad -\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2\} + o_p(1) \\
&= \frac{1}{(m!)^{d-1}} \sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d} \{\sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) + [\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) \\
&\quad -\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2\} + o_p(1) \\
&= \hat{\sigma}_n^2 \Big\{1 - \frac{1}{m^{d-1}} \sum_{\boldsymbol{i}} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n) - E^*(t_N^*(\hat{\boldsymbol{\theta}}_n))\Big\} \\
&\quad + \frac{1}{(m!)^{d-1}} \sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d} [\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 + o_p(1),
\end{aligned}$$

where $t_N^*(\hat{\boldsymbol{\theta}}_n) = \frac{1}{(1+\hat{g}^*)} \gamma_N^*(\hat{\boldsymbol{\theta}}_n)^T D_N^{*\,-1}(\hat{\boldsymbol{\theta}}_n) E_N^*(\hat{\boldsymbol{\theta}}_n) D_N^{*\,-1}(\hat{\boldsymbol{\theta}}_n) \gamma_N^*(\hat{\boldsymbol{\theta}}_n)$ and $\sum_{\boldsymbol{\pi}}$ is the summation of independent permutation over $\{0, 1, \ldots, m-1\}$.

The predictive distribution of normal prediction method has variance

$$\begin{aligned}
\sigma_2^{2*} &= E^*\{\sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^{*(u)}, \boldsymbol{y}_N^{*(u)}, \hat{\phi}_n) + o_p(1) \\
&= \frac{1}{(m!)^{d-1}} \sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d} \{\sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) + o_p(1) \\
&= \hat{\sigma}_n^2 \Big\{1 - \frac{1}{m^{d-1}} \sum_{\boldsymbol{i}} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n) - E^*(t_N^*(\hat{\boldsymbol{\theta}}_n))\Big\} + o_p(1)
\end{aligned}$$

Under condition A.3, we have $t_N^*(\cdot) \to 0$ prob-$P_{N,\omega}^*$ prob-P. Then the result follows. Comparing the predictive variance under both methods, it is straightforward to show that

$$\sigma_1^{2*} - \sigma_2^{2*} = E^*[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^{*(u)}, \boldsymbol{y}_N^{*(u)}, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 + o_p(1)$$

i.e. $P(\sigma_1^{2*} \geq \sigma_2^{2*}) \to 1$ as $n \to \infty$ $\square$

1.7.4 Proof of Theorem 3

Using the same technique in proof of Theorem 2, it is easy to show that the variance of the plug-in predictive distribution can be written as

$$\hat{\sigma}_n^2 \Big\{ 1 - \sum_{\boldsymbol{i}} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n) - \frac{1}{(1+g)} \gamma_n(\hat{\boldsymbol{\theta}}_n)^T D_n^{-1}(\hat{\boldsymbol{\theta}}_n) E_n(\hat{\boldsymbol{\theta}}_n) D_n^{-1} \gamma_n(\hat{\boldsymbol{\theta}}_n) \Big\},$$

Under condition A.3, we have $t_n(\cdot) = \frac{1}{(1+g)} \gamma_n(\hat{\boldsymbol{\theta}}_n)^T D_n^{-1}(\hat{\boldsymbol{\theta}}_n) E_n(\hat{\boldsymbol{\theta}}_n) D_n^{-1} \gamma_n(\hat{\boldsymbol{\theta}}_n) \to 0$ in P. Deducting the predictive variance σ_1^{2*} and σ_2^{2*} calculated in Theorem 2, the result follows immediately. $\square$

1.7.5 Proof of Theorem 4

Under the regularity assumptions given in Appendix, we compare the predictive variance on both in-sample and out-of-sample case under direct density approach and normal approximation approach. For the direct density approach, denote the variance within the sampled data by $\sigma_1^{2*(I)}$ and the variance for out-of-sample by $\sigma_1^{2*(O)}$. Similarly, we have $\sigma_2^{2*(I)}$ and $\sigma_2^{2*(O)}$ for the normal approximation method. We predict y at a given value x_{n+1}.

In one single m-LHD subsamples $(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*)$,
when x_{n+1} is within $(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*)$, by the interpolation property of Gaussian Process Model, we have

$$\hat{\sigma}_{n+1}^2 = 0$$

when x_{n+1} is out of $(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*)$, according to proof of Theorem 2 we have

$$\begin{aligned}
\hat{\sigma}_{n+1}^2 &= \sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) + [\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 \\
&\quad + o_p(1).
\end{aligned}$$

Under the regularity assumptions given in Appendix

$$\begin{aligned}
\sigma_1^{2*(I)} &= (1 - \frac{1}{m^{d-1}})\hat{\sigma}_n^2\Big\{1 - \frac{1}{m^{d-1}}\sum_{\boldsymbol{i}} \gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n) - E^*(t_N^*(\hat{\boldsymbol{\theta}}_n))\Big\} \\
&\quad + (1 - \frac{1}{m^{d-1}})\frac{1}{(m!)^{d-1}}\sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 \\
&\quad + \frac{1}{m^{d-1}} * 0 + o_p(1) \\
&= (1 - \frac{1}{m^{d-1}})\Big\{\hat{\sigma}_n^2[1 - \frac{1}{m^{d-1}}\sum_{\boldsymbol{i}} \gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)] \\
&\quad + \frac{1}{(m!)^{d-1}}\sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2\Big\} + o_p(1)
\end{aligned}$$

Under the normal approximation approach, similarly, when x_{n+1} is within $(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*)$, by property of interpolation of Gaussian Process Model,

$$\hat{\sigma}_{n+1}^2 = 0$$

when x_{n+1} is out of $(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*)$, according to proof of Theorem 2 we have

$$\hat{\sigma}_{n+1}^2 = \sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) + o_p(1)$$

Under the regularity assumptions given in Appendix

$$
\begin{aligned}
\sigma_2^{2*(I)} &= (1-\frac{1}{m^{d-1}})\hat{\sigma}_n^2\Big\{1-\frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)-E^*(t_N^*(\hat{\boldsymbol{\theta}}_n))\Big\} \\
&\quad +\frac{1}{m^{d-1}}*0+o_p(1) \\
&= (1-\frac{1}{m^{d-1}})\hat{\sigma}_n^2[1-\frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)]+o_p(1)
\end{aligned}
$$

According to Theorem 2,

$$
\begin{aligned}
\sigma_1^{2*(O)} &= \hat{\sigma}_n^2\Big\{1-\frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)\Big\} \\
&\quad +\frac{1}{(m!)^{d-1}}\sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*,\boldsymbol{y}_N^*,\hat{\phi}_n)-\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n,\boldsymbol{y}_n,\hat{\phi}_n)]^2+o_p(1)
\end{aligned}
$$

and

$$
\sigma_2^{2*(O)} = \hat{\sigma}_n^2\Big\{1-\frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)+o_p(1)
$$

To compare the in-sample and out-of-sample predictive variance, simply take the difference under the corresponding approach and the result follows immediately, we have

$$
\sigma_2^{2*(O)}-\ \sigma_2^{2*(I)}=\frac{\hat{\sigma}_n^2}{m^{d-1}}[1-\frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)]+o_p(1)
$$

i.e. $P(\ \sigma_2^{2*(O)}\geq\ \sigma_2^{2*(I)})\to 1$ as $n\to\infty$

and

$$\begin{aligned}
\sigma_1^{2*(O)} - \ \sigma_1^{2*(I)} &= \frac{1}{m^{d-1}}\Big\{\hat{\sigma}_n^2[1 - \frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)] \\
&\quad + \frac{1}{(m!)^{d-1}}\sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2\Big\} \\
&\quad + o_p(1) \\
&= \sigma_2^{2*(O)} - \ \sigma_2^{2*(I)} + o_p(1) \\
&\quad + (mm!)^{1-d}\sum_{\boldsymbol{\pi}}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 \\
&\geqslant 0
\end{aligned}$$

i.e. $P(\ \sigma_1^{2*(O)} \geq \ \sigma_1^{2*(I)}) \to 1$ as $n \to \infty$

$\square$

Chapter 2
Emulating Antarctic Ice-sheet Contributions to Sea Level Rise

2.1 Introduction

Sea level rise is expected to impact millions of people in coastal communities in the coming centuries (Strauss et al., 2012). But there are large uncertainties in how sea levels will change locally and globally, due primarily to a) uncertainties in Antarctic ice-sheet dynamics and b) the breadth of possible human choices with respect to emissions (Kopp et al., 2017). Modeling the contributions of the Antarctic ice-sheet (AIS) to sea level rise is challenging, in part because the physical mechanisms for ice-sheet instabilities are not well constrained. Furthermore, simulations of ice-sheet evolution in response to climate forcing are computationally expensive to run, limiting the range and quantity of experiments that may be performed.

Previously, expert elicitation has been used to describe the range of possible sea level contributions from the AIS (Bamber and Aspinall, 2013), but these judgments (and their accompanying probabilities) are limited because they are not explictly tied to dynamical or physical processes. But without a broad range of many representative model experiments, probability distributions of sea level contributions from Antarctica will either be based exclusively on discrete (non-continuous) outcomes (and hence will be less generalizable), or expert judgments will remain the primary descriptors for AIS uncertainties. In this study, we develop an emulator of an ice-sheet simulator to make progress overcoming these limitations.

The role of AIS contributions in sea level rise is not well constrained. Spatially complete modern observations of Antarctic mass changes (and hence sea level contributions) are extremely limited, with the satellite record dating back to 1993 (Team, 2018). Instead, it is necessary to turn to the past for analogues of sea levels which were significantly influenced by the Antarctic ice-sheet. In warm paleo-climates, in particular the "last interglacial" (LIG, sometimes also referred to as Marine Isotope Stage 5e; $\sim$130,000 to 115,000 years ago, (Kukla et al., 2002; of PAGES, 2016)) and the Pliocene ($\sim$3 million years ago, (Rovere et al., 2014; Pagani et al., 2009; Pollard et al., 2018)), global mean temperatures were up to 3 degrees Celsius warmer than present and peak global mean sea levels were potentially up to 9 or 30 meters higher than they are today, respectively. Although these geological records (and their constraints on sea level processes) remain rather uncertain (Horton et al., 2018), paleo-climate data (proxy reconstructions of sea level) suggests that only by invoking the role of the AIS can one account for the inferred large changes in sea level, because the other contributors (oceanic thermal expansion, mountain glaciers, and the Greenland ice-sheet) account for only part of the highstand estimates (Dutton et al., 2015). Until very recently, ice-sheet models were unable to simulate these expected large AIS contributions (Pollard et al., 2015).

A study by DeConto and Pollard (DeConto and Pollard, 2016) presented some of the first simulations of AIS contributions to global mean sea levels consistent with these likely contributions from the LIG and Pliocene periods. Their results showed that the AIS's implied role during these periods could be simulated once they had accounted for two physical processes: the hydrofracturing of ice-shelves and the mechanical failure of unstable ice cliffs. These processes were historically under-explored and unincorporated into ice-sheet models, but their inclusion in the DeConto and Pollard modeling framework provided simulations more consistent with geological evidence than previous work. Furthermore, future simulations including these processes projected median 21st century AIS contributions significantly higher than those previously published (Golledge et al., 2015; Ritz et al., 2015).

Although these newly integrated physical processes were parameterized based on modern observations (e.g., the retreat rate of the Jakobshavn Isbrae Glacier, $\sim$12 km/yr), the parameter values in (DeConto and Pollard, 2016) are uncertain, and the simulations were meant to be illustrative, "represent[ing] an envelope of possible outcomes." In fact, a wide range of solutions with different parameter calibrations are consistent with geological analogues (in part because the range of peak sea levels during these periods themselves have large uncertainties, cf. Figure 2.2.2). Accordingly, to explore projections of future AIS contributions to sea level rise, simulations must be computed over a broad set of these (loosely constrained) parameter values. This permits a description of some of the current uncertainties in the ice-sheet/climate coupled system. To carefully calibrate these model parameters conditional on modern and geological observations, it is necessary to obtain a full suite of simulations covering their possible ranges and accounting for their interactions.

Furthermore, it is important to incorporate scenario-dependant uncertainties from the climate system itself. Figure 2.1.1 (bottom) describes the goal framework for our projections of AIS contributions to sea level. There is a coupling between a representative climate system simulator (here, a regional climate model with global climate model boundary conditions) and an ice-sheet model which provides projections of sea level equivalent as a function of input user-chosen ice-sheet model parameters and climate emission scenarios. However, end-to-end (from climate model forcing input to ice-sheet simulation

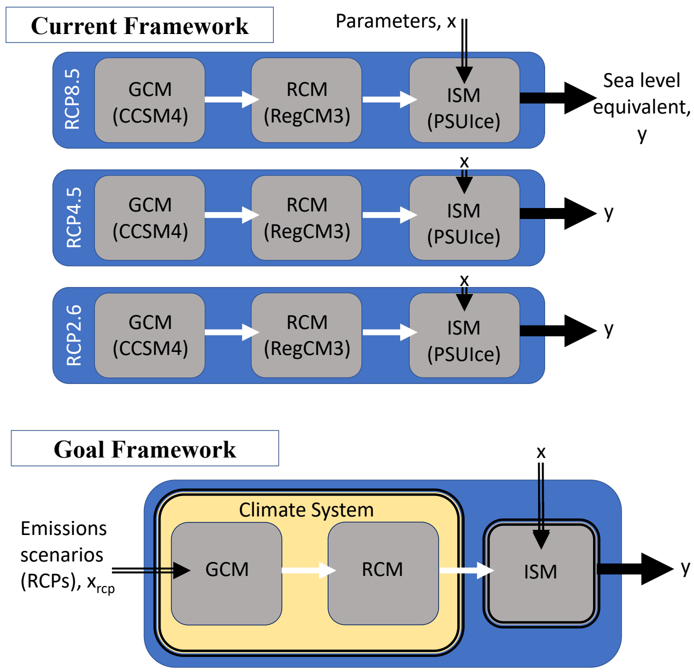


Figure 2.1.1: (top) Our current methodological framework for emulation, targeting at capturing the response of the ice-sheet model (PSUice) to changes in user chosen parameters endogenous in the ISM. In this formulation, the coupled global/regional/ice-sheet model system is treated as a black box for fixed climate emissions scenarios (i.e. RCP2.6, 4.5, or 8.5), and the emulator is developed to map inputs (x) to outputs (y) separately for each scenario. (bottom) A proposed framework which is an eventual goal of this research. Under this system, in addition to model parameters, emissions scenario would also be an emulator input (x_{rcp}). Climate system components may be exchanged for a global energy balance model or earth system model with appropriate inputs for the ISM. Black double-line arrows are the inputs, double-line boxes are components which may be substituted, the bold black arrow is the output, and white arrows represent the input/output communication between components.

output), this framework takes more than one month to complete for a single run, which is prohibitively expensive for obtaining a broad suite of simulations.

As an alternative, it is possible to construct and train a statistical emulator of this framework (using a discrete ensemble of ice-sheet model simulations) which is capable of producing solutions that are nearly continuous and can be computed rapidly (compared with the simulation computing time). Statistical emulation additionally produces solutions (and associated uncertainties) for previously untested model parameter values or emissions scenarios.

In this study, we adopt an approach describing the relationships between ice-sheets and model parameters to emulate over the ice-sheet model's parametric uncertainty. We define and fit a spline-based statistical emulator whose parameters are assigned Gaussian-process prior distributions. As a step towards our goal framework, we develop this spline-based emulation with simulations from a single fixed anthropogenic emissions scenario, but over a portion of the ice-sheet model's parametric range.

Ideally, a full treatment of the deep uncertainties associated with AIS contributions to sea level is needed (Bakker et al., 2017). But such a treatment requires simulations over all uncertain model parameters, ice-sheet representations among different simulators, and potential contributions beyond what may be currently simulated; such a treatment is beyond the scope of this study (see discussion on ongoing work in section 2.2). Instead, we develop emulation techniques which may be generalized and adaptable to other simulator parameter sets or other ice-sheet models. In other words, our methods are designed so that the components in Fig. 2.1.1 are modular.

In section 2.2, we describe and explore the data used in this study, including the ice-sheet simulations we intend to emulate. In section 2.3, we introduce the spline-based emulation methodology and present its results both in modeling and prediction. Finally in section 2.4, we conclude with a discussion of pros and cons of the spline-based approach and proposed next steps.

2.2 Data: Ice-Sheet Physics and Model Simulations

We emulate an ice-sheet simulator: the Pennsylvania State University Ice-Sheet model (PSUice), developed by David Pollard and Rob DeConto and described in (Pollard and Deconto, 2012) and (Pollard et al., 2015). PSUice simulations have been used in several studies of ice-sheet contributions to past and future sea level (Pollard et al., 2015; DeConto and Pollard, 2016; Pollard et al., 2017; Kopp et al., 2017; Pollard et al., 2018).

There are four ensembles (courtesy of Rob DeConto) of ice-sheet evolution computed with PSUice: three Representative Concentration Pathway simulations of future ice-sheet projections (RCP2.6, RCP4.5, RCP8.5, see (IPCC, 2013)) over 1950-2500, and a last interglacial (LIG) simulation of the ice-sheet at its peak retreat $\sim$125 thousand years ago (ka). Each ensemble has output of changes in the ice-sheet volume (in units of meters of global mean sea level equivalent, directly related to changes in ice-sheet volume above flotation) over time for the total Antarctic ice-sheet (TAIS), the East Antarctic ice-sheet (EAIS) and West Antarctic ice-sheet (WAIS). The boundary conditions for each PSUice ensemble are provided by the Regional Climate Model 3 (RegCM3, (Pal et al., 2007)). Details on the one-way coupling between climate models and the ice-sheet model are provided in (DeConto and Pollard, 2016).

Although changes in the total AIS drive global mean sea levels, WAIS and EAIS are physically distinct and have differing gravitational, rotational and deformational impacts on local sea level (Gomez et al., 2010; Mitrovica et al., 2011, 2018). It is therefore useful to consider their differential responses to climate warming. Much of WAIS lies below sea level, with buttressing ice-shelves that extend over the ocean. The ice-sheet grounding lines in direct contact with the ocean sit precariously on the edge of reverse sloping beds, where bedrock elevations drop off into the WAIS interior. The rate of ice-flux across the grounding line (out of the ice-sheet and into the ocean) is proportional to the height of the ice-sheet at the grounding line. If ocean temperatures warm (e.g., through climate change), melt can occur at the grounding line, causing it to retreat. The retreated grounding line then has a greater height, and the retreat rate again increases in a positive feedback loop known as Marine Ice Sheet Instability (e.g. (Weertman, 1974; Schoof, 2007)). The topographical configuration of WAIS means it is particularly susceptible to Marine Ice Sheet Instability, making it the primary focus of many studies of AIS contributions to sea level rise (e.g., (Mercer, 1978; Vaughan, 2008; Bamber et al., 2009; Pollard and DeConto, 2009; Joughin et al., 2014)).

In contrast, much of the EAIS is grounded above sea level and is more resilient to Marine Ice Sheet Instability. However, EAIS also stores significantly more ice-mass than WAIS ($\sim$53-60 meters in EAIS compared with the $\sim$3-4 meters in WAIS, (Bamber et al., 2009; Fretwell et al., 2013)). Therefore, EAIS instability has the potential to contribute significantly to sea level rise.

While we have discussed the distinctions between the EAIS and WAIS here as if they each evolve uniformly, in reality there are many more subdivided sectors of the AIS (e.g. see Extended Data Fig. 2 in (Team, 2018)) which have unique responses to climate forcing (these responses while distinct, may be correlated (Little et al., 2013)) and individual impacts on regional sea level. Although the simulated outputs provided in our current ensembles are only separated into WAIS and EAIS, a future goal of the emulation is to extend to more distinct sectors (which will require more apportioned outputs for emulator training).

The instability process which was newly-invoked in DeConto and Pollard (DeConto and Pollard, 2016) is the Marine Ice Cliff Instability. Ice-shelves sitting over marine sectors provide critical buttressing which reduces mass loss of the ice-sheet. But as atmospheric warming increases, surface meltwater and rainfall (transitioning from snow to rain in warmer climate) increase, pooling and penetrating into the ice-shelf as deep crevasses. These crevasses may deepen and fracture off, and as the buttressing ice-shelf is removed, mechanically unstable cliff faces may develop. When these cliff-faces structurally fail and the ice-sheet begins to retreat on a reverse sloping bed, a positive feedback loop begins (i.e. Marine Ice Cliff Instability). This instability can contribute greatly to sea level rise, as we will show. Modeled projections without ice-shelf fracturing and ice cliff collapse have markedly lower projected sea level contributions over the next 100 years than those which include them (Schlegel et al., 2018).

In PSUice, the Marine Ice Cliff Instability is primarily represented by two equations ((Pollard et al., 2015)) The first describes the horizontal ice-shelf calving rate C_{is} (in km/yr):

$$C_{is} = 3 * max(0, min(1, (d - d_c)/(1 - d_c))) \tag{2.1}$$

where, d_c is the critical fractional depth of the ice-sheet thickness (set to 0.75 in PSUice), and d is the ratio of crevasse depth to ice-sheet thickness. The depth of crevassing is related to a number of physical processes, including deepening from surface meltwater (see below). If ice-sheet crevasses deepen to exceed the critical depth, calving and ice-shelf retreat is initiated with retreat up to 3 km/yr.

Likewise, the horizontal cliff retreat rate, C_{ic} (in km/yr), in PSUice is:

$$C_{ic} = CLIFVMAX * max(0, min(1, (h_m - h_c)/20)) \tag{2.2}$$

where, h_c is the critical height of an ice-cliff before it becomes structurally unstable (set to

100 meters in PSUice), and h_m is the ice thickness above the water line modified by back-stresses and wet-crevasse deepening. If an ice cliff develops and its modified thickness exceeds the critical height, the cliff collapses, and the initiated retreat can accelerate up to the level of the $CLIFVMAX$ variable.

In our ensemble, the PSUice model configuration is varied over two-dimensional parameter space, where each parameter takes on 14 discrete values:

- ***CREVLIQ*** is the sensitivity of hydrofracturing in the model to surface liquid (i.e. from rain and meltwater). As this parameter increases, crevasses in the ice-sheet (d in Eqn. 2.1) deepen for given surface liquid accumulation rate, which increases the chance of the fracturing and removal of buttressing ice-shelves. *CREVLIQ* varies in the ensembles with values: 0, 15, 30, ..., 195 $\frac{m}{(m/yr)^2}$.

- ***CLIFVMAX*** is the maximum rate of horizontal cliff wastage once an ice-cliff becomes mechanically unstable (in Eqn. 2.2). As this parameter increases, structurally unsound ice-cliffs are permitted to retreat at a rate up to the increased bound. *CLIFVMAX* varies in the ensembles with values: 0, 1, 2, ..., 13 km/yr.

For each scenario (RCPs or LIG), there are a total of $14 \times 14 = 196$ ensemble members (every combination of the discrete configurations of *CREVLIQ* and *CLIFVMAX*). We note that when *CLIFVMAX*$= 0$, ice cliffs are not permitted to retreat even when they should structurally fail; in this case, the Marine Ice Cliff Instability mechanism is “turned off” in the model simulations.

Ensemble simulations for the LIG scenario and the RCP8.5 scenario are shown in Figure 2.2.2. The LIG ensemble was designed as an equilibrium scenario to represent the peak of ice-sheet retreat during the period, so we consider only the final value in our analyses (labeled 125ka). Comparing with the estimates of total AIS contributions to sea

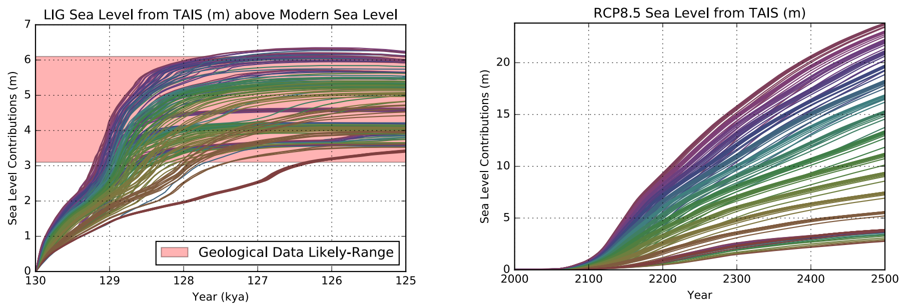


Figure 2.2.1: Timeseries ensembles of ice-volume lost from the total Antarctic ice-sheet (in meters of sea level equivalent) for the LIG (left) and RCP8.5 (right) scenarios. Runs are color-coded spanning *CLIFVMAX*$=0$ (yellow) to *CLIFVMAX*$=13$ (purple). Red shading in the LIG figure is the geological likely-range of sea level contributions from total AIS (Dutton et al., 2015).

level from geological data (Dutton et al., 2015) shown in the red shading, we find that all but a few of the highest model ensemble members fall within the estimated historical range.

Sea level contributions from the total AIS increase as the values of *CLIFVMAX* or *CREVLIQ* increase (Figure 2.2.2; driven largely by Marine Ice Cliff Instability). The two scenarios produce distinct sensitivities to the parameter values: *CLIFVMAX* dominates sea level rise in 2100 under the RCP8.5 forcing, whereas both *CLIFVMAX* or *CREVLIQ* are important in the LIG scenario. Critically, there are interactions between *CLIFVMAX* and *CREVLIQ*; for instance, without a sufficient maximum cliff retreat rate (the leftmost columns of Fig. 2.2.2), hydrofracturing of buttressing ice-shelves does not significantly increase sea levels.

The differences between EAIS and WAIS for the RCP8.5 scenario are illustrated in the broad spread (here represented by the 5-95th quantiles) of the ice-sheet ensembles shown in Figure 2.2.3. WAIS consistently loses mass and has a relatively small range in 2100, suggesting that instabilities are expected during the 21st century across the entire suite of parameter values. In contrast, some simulations of EAIS actually have *negative* sea level

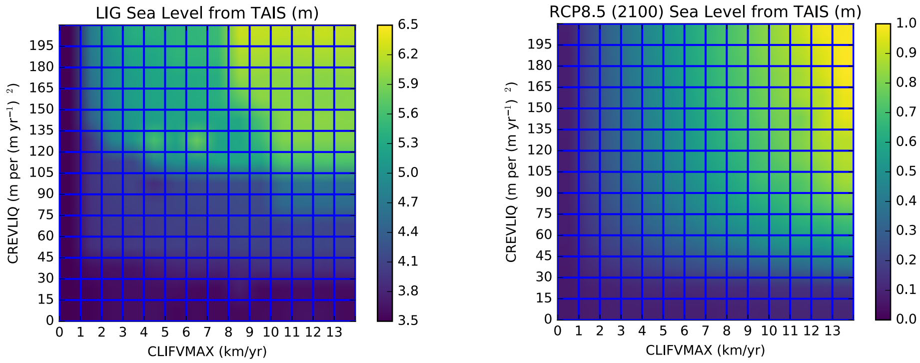


Figure 2.2.2: Contoured ensemble sea level equivalent contributions (m) from the total AIS for the LIG scenario at peak retreat (∼125ka, left) and for the RCP8.5 scenario in 2100 (right), as a function of model parameters *CREVLIQ* (y-axis) and *CLIFVMAX* (x-axis).

contributions through 2100, which is indicative of ice-sheet mass gains from increased accumulation of snow (related to warmer atmospheric temperatures, see (Medley et al., 2018) and references therein).

One limitation of the ensembles we train the emulator on is that only two PSUice parameters are varied, whereas there are >60 parameters that can be manipulated (Rob Deconto, *personal communication*). In particular two such parameters, the model's coefficient of sub-ice-shelf oceanic melt (referred to as OCFAC in (DeConto and Pollard, 2016)) and the isostatic rebound relaxation rate (referred to as τ, usually around 3000 years, (Pollard et al., 2015)), could dramatically influence sea level contributions from the AIS. Complete descriptions of these (and all) model parameterizations are found in (Pollard et al., 2015). Work by David Pollard and coauthors is ongoing to explore parameter/variable selection for PSUice; a long-term goal of this research is to integrate our research with theirs. The solutions with parameters *CREVLIQ* and *CLIFVMAX*, however, represent a broad illustrative range of how the ice-sheet may respond to climate warming, and are therefore useful to emulate. Extension to a larger model parameter space is left to future work.

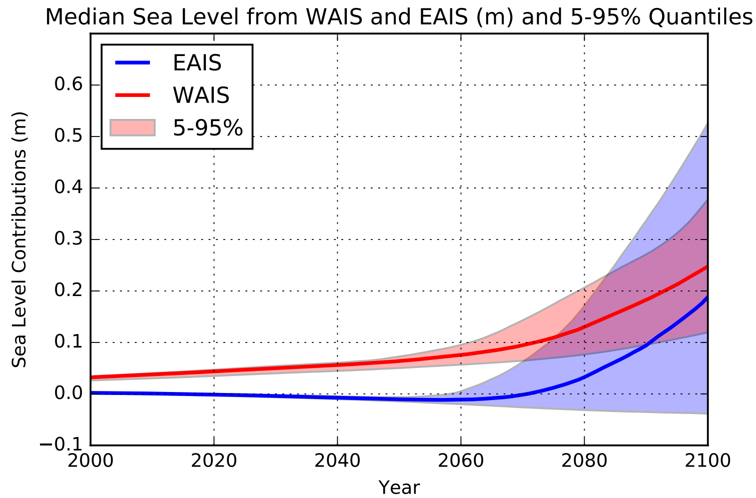


Figure 2.2.3: The median and 5-95th quantiles of the ensemble projections of sea level contributions from the Western (red) and Eastern (blue) Antarctic ice-sheets for the RCP8.5 scenario during the 21st century.

2.3 Spline-based emulation

2.3.1 Background and Motivation

To build a statistical emulator for the ice-sheet model and give prediction uncertainty at the same time, some important properties of this data set need to be addressed. Firstly, sea level contribution(output) has a nonlinear relationship with *CREVLIQ*, *CLIFVMAX*, and *time* (input). Secondly, sea level contribution behaves quite differently in relation to *time* than to the ice-sheet parameters, *CREVLIQ* and *CLIFVMAX*; i.e., sea level rise is smoother in *time* than in the ice-sheet model parameters. Classical models are not directly applicable to this data set. For example, polynomial regression of high order, as a method to model nonlinearity, can capture some portion of variance in the output variable; however, it has obvious disadvantages, including that it does not have enough flexibility to capture the full trend of the data. Moreover, polynomial regression can be highly sensitive to a single influential data point, and performs poorly for extrapolation. Another method that is widely used in computer experiments, due to its flexibility and uncertainty quantification for prediction, is Gaussian Process (GP) regression. However, one drawback to applying GP regression in three dimensions with a large sample size is the computational expense. In addition, the assumption of full Gaussian Process might be too strong here. As a result, we need a model in between, which can be flexible enough to capture the trend, while computationally feasible at the same time.

Spline regression is a powerful tool for modeling the nonlinear relationship between input and output(Marsh and Cormier, 2001). When using splines, the predictor domain is subdivided by a set of knots, and regression curves are fitted piecewise between each adjacent knot to join smoothly (continuously). One advantage of using splines is the balance between the complexity of a given model and its goodness of fit(Wegman and Wright, 1983; Wold, 1974). Splines additionally allow the *time* dimension to be treated differently from *CREVLIQ* and *CLIFVMAX*. Cubic splines (i.e., a piecewise polynomial of degree 3) are the most commonly used as it is quite smooth and cost effective(Wang and Wu, 2017).

A graphical illustration of spline regression with a single set of values for *CREVLIQ* and *CLIFVMAX* is shown in Figure 2.3.1, where the plot on the left fits a model with 1 knot and the plot on the right with 2 knots. Spline regression works quite well for a single set of parameters, *CREVLIQ* and *CLIFVMAX*. However, one drawback of simple spline regression is that it does not provide prediction uncertainties. At the same time, we are modelling a single set of values for *CREVLIQ* and *CLIFVMAX* separately. To overcome these two limitations, we propose to impose a Gaussian Process prior on the coefficients of the spline regression. The use of a covariance function within the Gaussian Process enables sharing of information through the high correlation of hyperparameters (parameters of the spline regression model) for nearby ice-sheet parameters. Conversely,

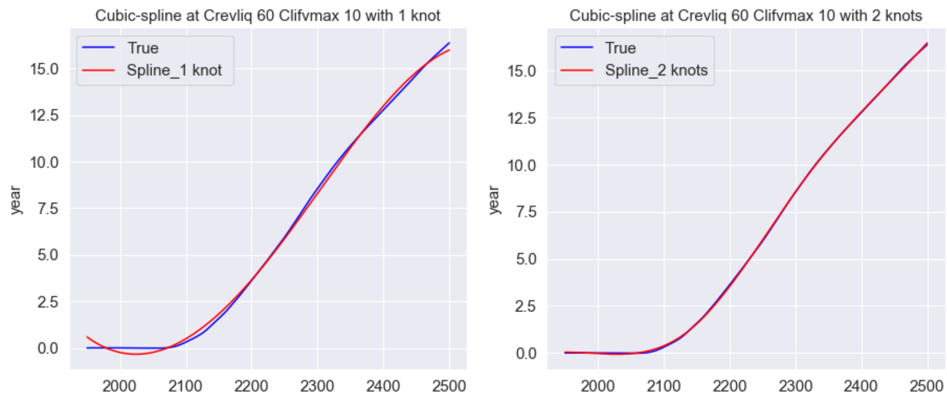


Figure 2.3.1: Cubic spline models at *CREVLIQ*=60 and *CLIFVMAX*=10 with one knot (left) and two knots (right)

as the distance increases between values of *CREVLIQ* and *CLIFVMAX*, the correlation in hyperparameter values diminishes (A.O'Hagan and J.F.C.Kingman, 1978).

To simplify the modelling and illustrate this approach, we first reduce the size of the data set by averaging the temporal data over 19-year periods and fixing *CREVLIQ* = 60. Within the framework of this spline method with GP priors on hyperparameters, we can control the trade-off between fidelity to the data and roughness of the function estimate.

2.3.2 Methodology

A cubic spline model takes the following form:

$$y(x,t) = \beta_0(x) + \beta_1(x)t + \beta_2(x)t^2 + \beta_3(x)t^3 + \sum_{k=1}^{k=K} \beta_{k+3}h(t,\xi_k) + \epsilon \tag{2.3}$$

$$= \boldsymbol{f}(t)^T \boldsymbol{\beta}(x) + \epsilon \tag{2.4}$$

where K denote the number of knots, x denotes *CLIFVMAX* values, $\beta_i, i = 0, 1, ...(K+3)$ are the parameter coefficients of the corresponding basis functions and t is centered and normalized by its standard deviation. Gaussian uncertainty, $\epsilon, \sim N(0, \sigma^2)$ is assumed. The piecewise functions between knots are defined as:

$$h(t, \xi) = (t - \xi)_+^3 = \begin{cases} (t - \xi)^3, & \text{if t} > \xi \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

Here, t time denotes time segments, averaged over 19 years and ξ represents each knot. A spline regression with $K = 4$ knots is applied to the reduced data set, where the years of the knots location ξ_1 to ξ_4 are year 2016, year 2054, year 2092 and year 2225. Knots are located denser in the early years because the discrepancy between the ice-sheet emulator and the simulator is less tolerable in 21st century. We allow different values of $\boldsymbol{\beta}$ to be appropriate for different x.

To ensure a single value of $\boldsymbol{\beta}$ provides adequate approximation to the curve locally as x varies, we assume $\boldsymbol{\beta}(x)$ and $\boldsymbol{\beta}(x^*)$ highly correlated when x and x^* are close. The belief in local stability can be translated into a suitable prior distribution of $\boldsymbol{\beta}(x)$. Here we assume $\boldsymbol{\beta}(x)$ follows a Gaussian Process (GP) with prior mean and covariance function defined as:

$$\mathbf{E}\boldsymbol{\beta}(x) = \boldsymbol{b_0}, \tag{2.6}$$

$$\mathbf{Cov}(\boldsymbol{\beta}(x), \boldsymbol{\beta}(x^*)) = \rho(|x - x^*|)\boldsymbol{B_0} \tag{2.7}$$

$\rho(d)$ (the correlation function) is a monotonically decreasing function with $d \geq 0$ and $\rho(0) = 1$. The more slowly $\rho(d)$ decreases to 0 the more stable $\boldsymbol{\beta}(x)$ is. $\boldsymbol{b_0}$ represents the prior mean function for $\boldsymbol{\beta}(x)$. Assuming separability of $\boldsymbol{\beta}(x)$, which gives us some computational advantages, $\boldsymbol{B_0}$ is the 8×8 diagonal variance-covariance matrix of $\boldsymbol{\beta}(x)$. If we assume a Matérn 3/2 correlation function, which gives fits with the similar degree of smoothness as the cubic splines, then $\rho(|x - x^*|) = \exp(-|x - x^*|/r) * (1 + |x - x^*|/r)$

where r is the range parameter of ρ.

Suppose we observe n values of the dependent variable, $y_1, ...y_n$, at $(x_1, t_1), (x_2, t_2), ...$ (x_n, t_n), $x_i = 1, 2, ...m, t_j = 1, 2, ...q$. The posterior distribution of $\boldsymbol{\beta}(x)$ is jointly normal with mean and variance

$$\boldsymbol{b}(x) = E(\boldsymbol{\beta}(x)|y_1, ..y_n, \boldsymbol{b_0}) = \boldsymbol{S}(x)^T \boldsymbol{A}^{-1} \boldsymbol{y} + \boldsymbol{Q}(x)^T \boldsymbol{b_0}(x), \tag{2.8}$$

$$\begin{aligned} \boldsymbol{B}(x, x^*) &= E(\{\boldsymbol{\beta}(x) - \boldsymbol{b}(x)\}\{\boldsymbol{\beta}(x^*) - \boldsymbol{b}(x^*)^T\}|y_1, ...y_n, \boldsymbol{b_0}) \\ &= \rho(|x - x^*|)\boldsymbol{B_0} - \boldsymbol{S}(x)^T \boldsymbol{A}^{-1} \boldsymbol{S}(x^*) \end{aligned} \tag{2.9}$$

where $\boldsymbol{Q}(x) = \boldsymbol{I}_{K+4} - \boldsymbol{G}^T \boldsymbol{A}^{-1} \boldsymbol{S}(x),\ \boldsymbol{y} = (y_1, ...y_n)$, K denotes the number of knots $\boldsymbol{S}(x)$ is the $n\times$ (K+4) matrix whose i^{th} row is $\rho(|x - x_i|)\boldsymbol{f}(t_i)^T \boldsymbol{B_0}$,

$$\boldsymbol{G}(t) = (\boldsymbol{f}(t_1)^T, ..., \boldsymbol{f}(t_n)^T), \boldsymbol{A} = \sigma^2 \boldsymbol{I_n} + \boldsymbol{C},$$

where $\boldsymbol{C}$ is $n \times n$ with $(i, j)^{th}$ elements $c_{i,j} = \rho(|x_i - x_j|)\boldsymbol{f}(t_i)^T \boldsymbol{B_0} \boldsymbol{f}(t_j)$

For any future value y at (x, t), the predictive distribution of y is

$$N(\boldsymbol{f}(t)^T \boldsymbol{b}(x), \sigma^2 + \boldsymbol{f}(t)^T \boldsymbol{B}(x, x)\boldsymbol{f}(t)) \tag{2.10}$$

2.3.3 Results

The results shown in this section, if not otherwise specified, are based on the data of EAIS in RCP8.5 introduced in 2.2. The building of the statistical emulator can be directly extended to the data in WAIS and other RCPs. Limited by space here, we do not show all the results for different scenarios but take the data of EAIS and RCP8.5 as an illustration example to show the performance of the proposed spline-based approach.

2.3.3.1 Estimation of priors b_0 and $\boldsymbol{B_0}$

The mean prior b_0 and diagonal covariance matrix prior $\boldsymbol{B_0}$ of $\boldsymbol{\beta}(x)$ can be estimated by likelihood-based methods. In this section, b_0 and $\boldsymbol{B_0}$ are estimated by maximizing the likelihood function. The estimated prior is given as follows:

$$\hat{b_0} = [-1.20, -2.31, -1.48, -0.32, 1.63, 3.54, -6.47, 1.38],$$
$$\hat{\boldsymbol{B_0}} = \mathrm{diag}\{8.97, 41.4, 20.9, 1.15, 3.22, 1.24, 0.15, 1.17e-3\}.$$

2.3.3.2 Estimated $\boldsymbol{\beta}(x)$ Coefficients

The posterior of $\boldsymbol{\beta}(x)$ at each x level can be obtained by Equations (2.8)-(2.9). The results of estimated $\boldsymbol{\beta}(x)$ are summarized both in Table 2.3.1 and Figure 2.3.2. In Table 2.3.1, the first column denotes the *CLIFVMAX* level. $\boldsymbol{\beta}(x)$ can be calculated for any *CLIFVMAX* level in the continuous space. Based on $\boldsymbol{\beta}(x)$, the prediction of sea level contribution at any time t can be provided by Equation 2.3. We show the estimation of $\boldsymbol{\beta}(x)$ at each discrete *CLIFVMAX* level in Table 2.3.1 along with its standard deviation. The estimation of $\boldsymbol{\beta}(x)$ at a continuous space of *CLIFVMAX* level is demonstrated in Figure 2.3.2 with its 95% confidence interval.

2.3.3.3 Contribution of each basis function in spline-based model

An example at *CLIFVMAX* = 12 is used to demonstrate the cumulative effect of each basis in the spline function in Figure 2.3.3. It provides an intuitive idea about how the spline-base approach works to approximate the target curve through each spline basis. β_0 to β_3 takes effect on the full time t horizon, while β_4 to β_7 only makes contribution beyond

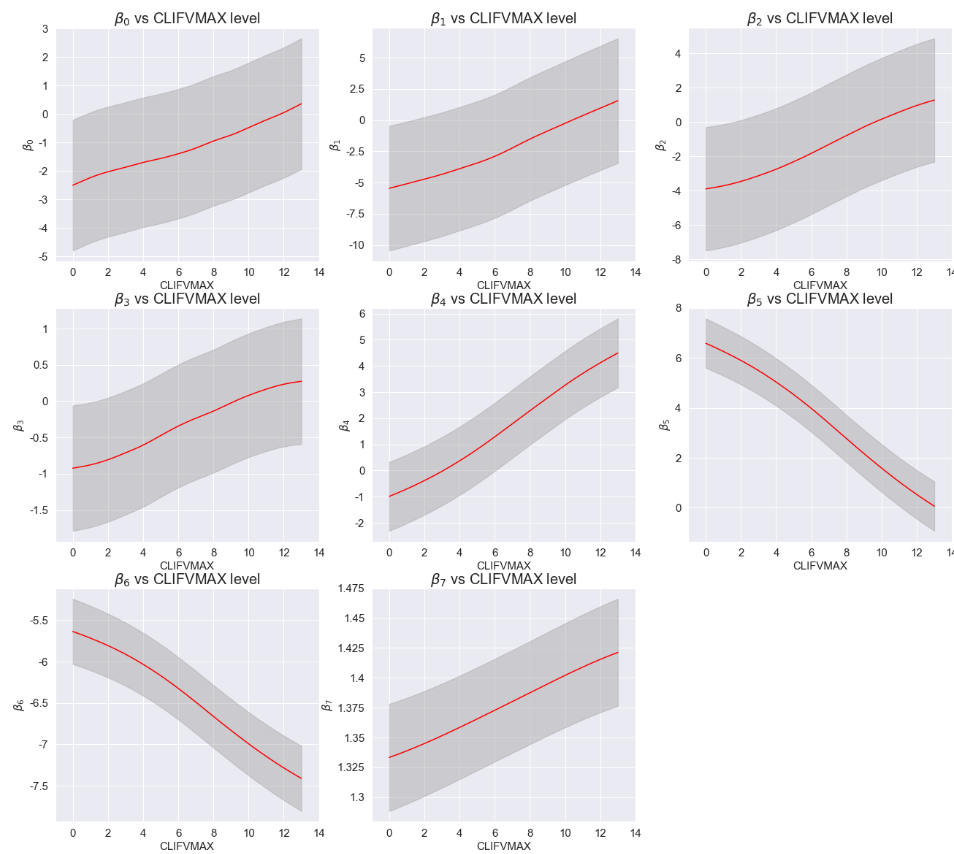


Figure 2.3.2: Posterior distributions of $\beta_i(x)$. From top left to bottom right, β(x) is presented in ascending order, from β_0(x) to β_7(x). Grey shading reflects 95% confidence interval (1.96 standard deviation away).

Table 2.3.1: Estimated $\boldsymbol{\beta}(x)$ coefficients at each *CLIFVMAX* level (standard deviation in parenthesis)). First column denotes *CLIFVMAX* level.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
0	-2.50(1.17)	-5.46(2.55)	-3.90(1.84)	-0.92(0.44)	-0.99(0.67)	6.59(0.50)	-5.64(0.20)	1.33(0.02)
1	-2.24(1.17)	-5.10(2.54)	-3.72(1.83)	-0.88(0.44)	-0.71(0.67)	6.26(0.50)	-5.72(0.20)	1.34(0.02)
2	-2.03(1.16)	-4.73(2.53)	-3.46(1.82)	-0.81(0.44)	-0.39(0.67)	5.90(0.49)	-5.81(0.20)	1.34(0.02)
3	-1.87(1.16)	-4.34(2.53)	-3.13(1.82)	-0.71(0.43)	-0.03(0.67)	5.49(0.49)	-5.91(0.19)	1.35(0.02)
4	-1.70(1.16)	-3.89(2.52)	-2.75(1.82)	-0.61(0.43)	0.37(0.67)	5.03(0.48)	-6.03(0.19)	1.36(0.02)
5	-1.56(1.16)	-3.44(2.52)	-2.31(1.81)	-0.47(0.43)	0.81(0.67)	4.53(0.48)	-6.17(0.19)	1.37(0.02)
6	-1.40(1.16)	-2.90(2.52)	-1.83(1.81)	-0.34(0.43)	1.29(0.67)	3.98(0.48)	-6.32(0.19)	1.37(0.02)
7	-1.20(1.16)	-2.23(2.52)	-1.30(1.81)	-0.24(0.43)	1.78(0.67)	3.39(0.48)	-6.49(0.19)	1.38(0.02)
8	-0.95(1.16)	-1.52(2.52)	-0.79(1.81)	-0.14(0.43)	2.29(0.67)	2.77(0.48)	-6.66(0.19)	1.39(0.02)
9	-0.74(1.16)	-0.88(2.52)	-0.28(1.82)	-0.02(0.43)	2.78(0.67)	2.17(0.48)	-6.83(0.19)	1.39(0.02)
10	-0.48(1.16)	-0.26(2.53)	0.17(1.82)	0.08(0.43)	3.26(0.67)	1.59(0.49)	-6.99(0.19)	1.40(0.02)
11	-0.21(1.16)	0.36(2.53)	0.58(1.82)	0.17(0.44)	3.71(0.67)	1.04(0.49)	-7.14(0.20)	1.41(0.02)
12	0.05(1.17)	0.96(2.54)	0.97(1.83)	0.23(0.44)	4.12(0.67)	0.53(0.50)	-7.28(0.20)	1.42(0.02)
13	0.36(1.17)	1.55(2.55)	1.28(1.84)	0.27(0.44)	4.49(0.67)	0.07(0.50)	-7.41(0.20)	1.42(0.02)

the location of its corresponding knot ξ_1 to ξ_4.

2.3.3.4 Prediction at each *CLIFVMAX* level

With the distribution of $\boldsymbol{\beta}(x)$, one can make prediction on sea level contribution at any value of *CLIFVMAX* and *time*. By doing Leave-one-out cross validation(LOOCV), the prediction at each *CLIFVMAX* is shown in figures 2.3.4a and 2.3.4b. To be more specific, the ”one” in LOOCV means to leave out an entire ice-sheet model run. From the prediction plots, it appears that the predictions are quite close to the true values from the ice-sheet model, with reasonable uncertainties. The comparison of the prediction and the corresponding actual value given by the simulation at each *CLIVMAX* level at year 2054, 2092 and 2491 are listed in the Table 2.3.2. Sea level contribution is averaged every 19 years denoted by the central year. (For example, sea level contribution at year 2054 is the sea level contribution averaged from year 2045 to year 2063.)

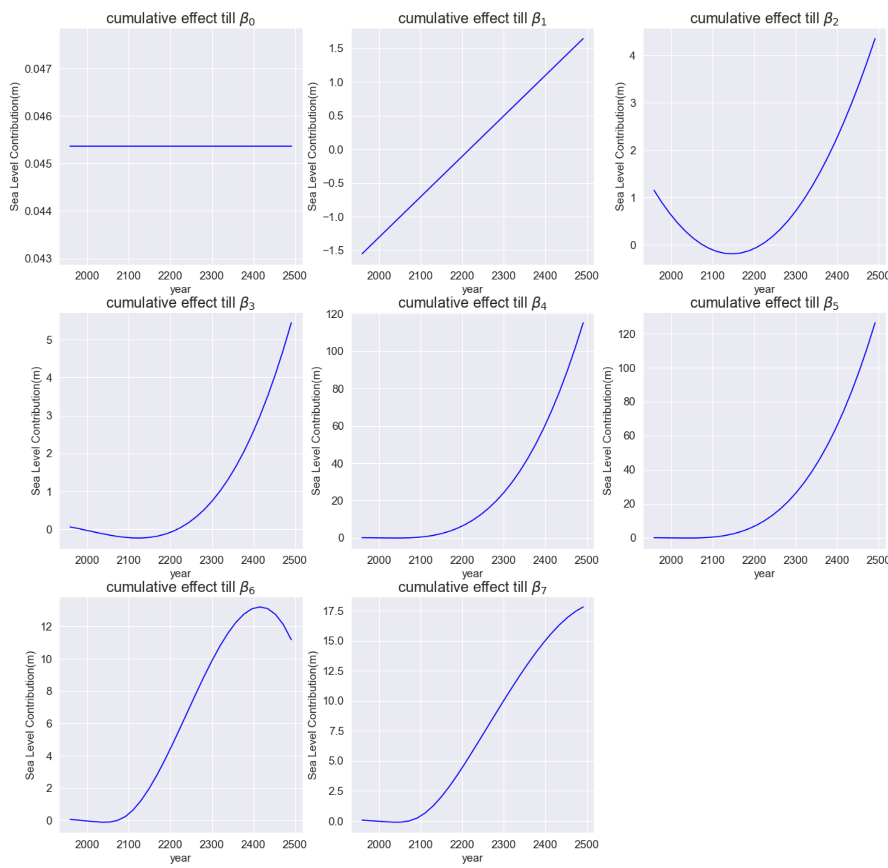


Figure 2.3.3: Cumulative effect of each basis function term in the spline-based model for one example sea-level curve at *CLIFVMAX* = 12. From the top left to the bottom right, $\beta_0(x)$, $\beta_1(x)t$,..., $\beta_7(x)(t-\xi_4)_+^3$ are added sequentially to the model. (Note that the scale of Sea Level Contribution is different in each subplot.)

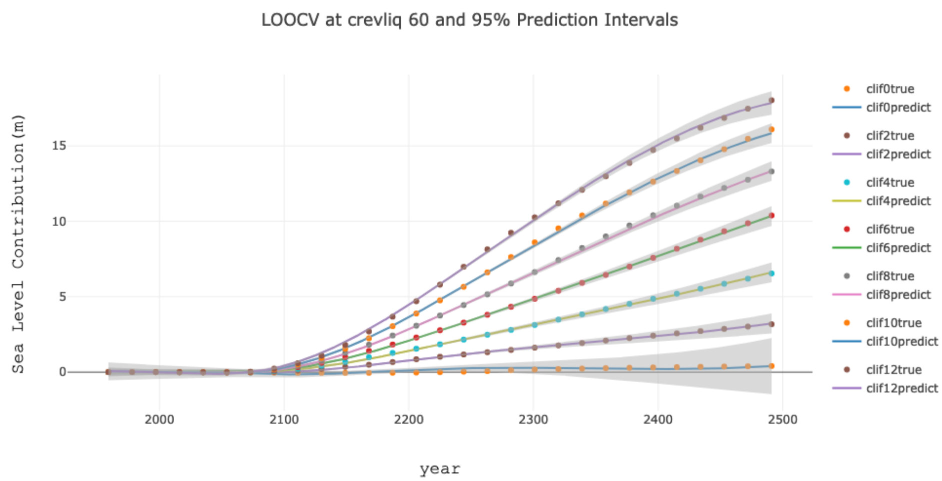
Table 2.3.2: Comparison of Prediction of Sea Level rise at each *CLIFVMAX* level (standard deviation in parenthesis)) and the actual value from simulation at specific years. First column denotes *CLIFVMAX* level.

CLIFVMAX	Year 2054		Year 2092		Year 2491	
	Actual(m)	Prediction(m)	Actual(m)	Prediction(m)	Actual(m)	Prediction(m)
0	-0.02	-0.02(0.026)	-0.04	-0.13(0.015)	0.41	0.40(0.897)
1	-0.02	0.02(0.005)	-0.00	-0.06(0.004)	1.65	1.81(0.158)
2	-0.01	0.05(0.004)	0.02	0.01(0.003)	3.18	3.23(0.114)
3	-0.01	0.06(0.004)	0.05	0.05(0.003)	4.85	4.85(0.112)
4	-0.01	0.07(0.004)	0.07	0.09(0.003)	6.54	6.61(0.112)
5	-0.01	0.07(0.004)	0.09	0.13(0.003)	8.50	8.50(0.112)
6	-0.01	0.05(0.004)	0.11	0.13(0.003)	10.39	10.35(0.112)
7	-0.01	0.02(0.004)	0.13	0.14(0.003)	11.96	12.08(0.112)
8	-0.01	-0.03(0.004)	0.15	0.13(0.003)	13.30	13.33(0.112)
9	-0.01	-0.06(0.004)	0.17	0.15(0.003)	14.58	14.67(0.112)
10	-0.01	-0.07(0.004)	0.19	0.18(0.003)	16.09	15.84(0.112)
11	-0.01	-0.08(0.004)	0.22	0.22(0.003)	17.18	17.03(0.114)
12	-0.01	-0.10(0.005)	0.24	0.26(0.004)	18.02	17.84(0.158)
13	-0.01	-0.12(0.026)	0.26	0.28(0.015)	18.81	18.48(0.897)

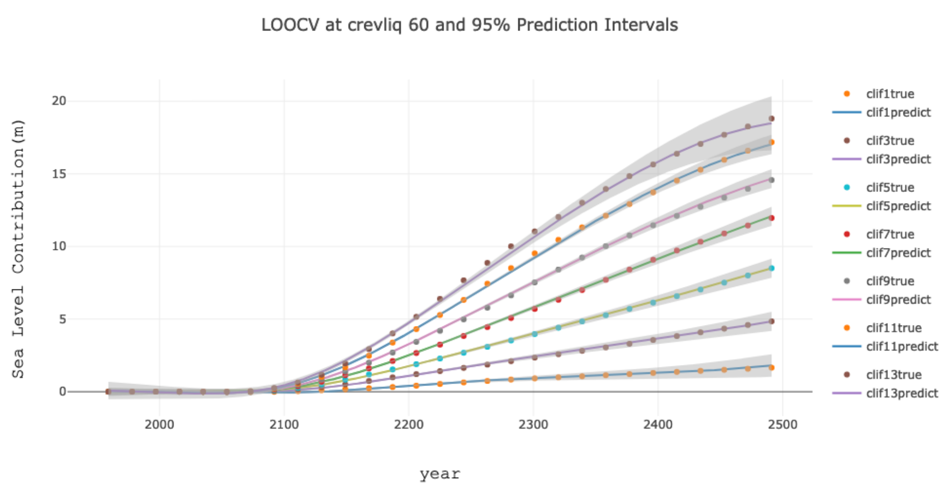
As expected, predictions at *CLIFVVMAX* $= 0$ and 14 have the widest prediction intervals among all the *CLIFVVMAX* value, because essentially extrapolation is implemented when making prediction at these two *CLIFVVMAX* levels. In the early years, the prediction uncertainty is quite small. As time passes by, the prediction uncertainty increases with the scale of sea level rise.

2.3.3.5 Residuals of full fitted model and prediction

The residual plot resulting from the spline-based fitted model is shown in Figure 2.3.5, in which the levels of *CLIFVMAX* are coded by color. We find that the majority of the sea-level signal is already explained by this model. Throughout the full time horizon from year 1950 to year 2500, the difference between the actual value and the fitted value is strictly within 30cm. The sea level rise contributions in 21st century draws the most attractions



(a)



(b)

Figure 2.3.4: Comparison of predictions using spline-based emulation and ice-sheet model using LOOCV. Grey shading reflects 95% confidence intervals (prediction uncertainty). (a)*CLIFVMAX* level is even. (b)*CLIFVMAX* level is odd.

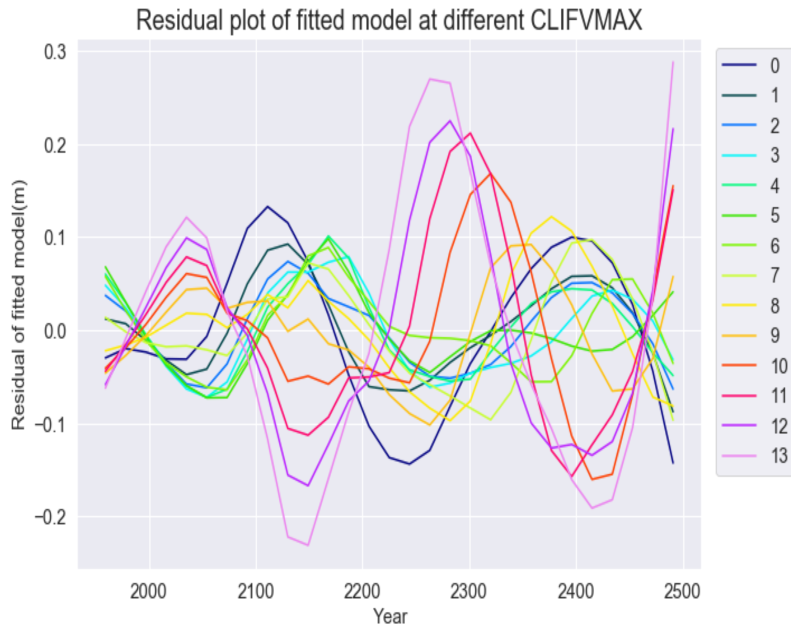


Figure 2.3.5: Residuals of the spline-based fitted model at each *CLIFVMAX* level. Each color matches one *CLIFVMAX* level displayed in the legend.

among researchers. From the residual plot, it shows that our model, in particular, works well in the early year with less than 10cm discrepancy most of the time.

2.3.3.6 Coverage ratio of prediction uncertainty

To assess the accuracy of prediction uncertainty, experimental coverage probability is used to estimate the nominal value calculated by dividing number of observations inside their Prediction Interval (PI) by the total number of observations. For each PI, the coverage ratio should be close to the given PI percentage. If more data fall within the PI than expected, the PI is too conservative. Whereas if fewer data fall within the PI, then then the PI is too aggressive. We find that the coverage ratio is 95.6% for the 99.7% PI (3 standard deviations above and below the predictive mean), 87.2% for the 95% PI (1.96 standard deviations from the predictive mean) and 69.0% for the 68% PI (1 standard deviation from the predictive mean). It is interesting to observe that the prediction uncertainty given by the spline-based model is almost consistent with the experimental coverage, which supports credibility of uncertainty quantification provided by the proposed method.

2.4 Discussion and Concluding Remarks

The spline-based statistical emulation for predicting outputs of the PSUice model addresses some of the limitations in the Antarctic ice-sheet simulator itself. It is continuous, enabling prediction at the entire range of *CLIFVMAX* for a single *CREVLIQ* value. It borrows information from close *CLIFVMAX* value through Gaussian Process prior to provide an statistical emulator with adequate approximation locally. Compared to full Gaussian Process model, it is parsimonious so that it brings computational advantages. The spline-based emulator provides a good fit with the given data set and credible prediction uncertainties for any prediction. With the current framework, it is easy to extend the spline-based approach to accommodate *CREVLIQ* as an input variable of the model, as well as more outputs from the climate model as inputs in an emulator of the ice-sheet model, if applicable, in the future.

Future work can be tried in the following directions. First, since the model fitting and prediction is more important before year 2100, except placing more knots between year 2000 and year 2100 as proposed in Section 2.3, applying a weighted loss (penalizing more on the discrepancy at the early year) when fitting the spline might also improve the accuracy of our model in 21st century. Second, Maximum Likelihood Estimation is used to give a rough estimation of the mean and covariance prior of the coefficients of the proposed model. More sophisticated likelihood-based method can be implemented to estimate the prior, as well as the range parameter in Matérn 3/2 correlation function. A fully Bayesian framework could also work here. In addition, there may be advantages to quantifying information from different emissions scenarios to inform ice-sheet changes to build a full model. Furthermore, information such as precipitation (provided by the climate model) could prove valuable for emulating changes in the ice-sheet volume. Modeling with these techniques will permit inquiries across any reasonable range of human emissions and accompanying climate changes, an important feature not yet available to the spline-based technique.

Bibliography

A.O'Hagan and J.F.C.Kingman (1978). Curve fitting and optimal design for prediction. *Wiley for the Royal Statistical Society*, 40(1):1–42.

Bakker, A. M., Wong, T. E., Ruckert, K. L., and Keller, K. (2017). Sea-level projections representing the deeply uncertain contribution of the West Antarctic ice sheet. *Scientific Reports*, 7(1):1–7.

Bamber, J. L. and Aspinall, W. P. (2013). An expert judgement assessment of future sea level rise from the ice sheets. *Nature Climate Change*, 3(4):424–427.

Bamber, J. L., Riva, R. E. M., Vermeersen, B. L. A., and LeBrocq, A. M. (2009). Reassessment of the {P}otential {S}ea-{L}evel {R}ise from a {C}ollapse of the {W}est {A}ntarctic {I}ce {S}heet. *Science*, 324(5929):901–903.

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

Bouttes, N., Gregory, J. M., and Lowe, J. a. (2013). The reversibility of sea level rise. *Journal of Climate*, 26(8):2502–2513.

Castrillon-Candás, J. E., Genton, M. G., and Yokota, R. (2016). Multi-level restricted maximum likelihood covariance estimation and kriging for large non-gridded spatial datasets. *Spatial Statistics*, 18:105–124.

Castrilln-Cands, J. E., Genton, M. G., and Yokota, R. (2016). Multi-level restricted maximum likelihood covariance estimation and kriging for large non-gridded spatial datasets. *Spatial Statistics*, 18:105 – 124. Spatial Statistics Avignon: Emerging Patterns.

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

Chu, T., Zhu, J., Wang, H., et al. (2011). Penalized maximum likelihood estimation and variable selection in geostatistics. *The Annals of Statistics*, 39(5):2607–2625.

Cressie, N. (1992). Statistics for spatial data. *Terra Nova*, 4(5):613–617.

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

DeConto, R. M. and Pollard, D. (2016). Contribution of Antarctica to past and future sea-level rise. *Nature*, 531(7596):591–597.

Diciccio, T. and Efron, B. (1992). More accurate confidence intervals in exponential families. *Biometrika*, 79(2):231–245.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical science*, pages 189–212.

Draguljić, D., Santner, T. J., and Dean, A. M. (2012). Noncollapsing space-filling designs for bounded nonrectangular regions. *Technometrics*, 54(2):169–178.

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

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, pages 569–593. Springer.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Fang, K.-T., Li, R., and Sudjianto, A. (2005). *Design and modeling for computer experiments*. Chapman and Hall/CRC.

for Atmospheric Research (UCAR), U. C. NCAR Climate Data Gateway, Search page. https://www.earthsystemgrid.org/search.html. (accessed: 2018-11-18).

Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J. L., Barrand, N. E., Bell, R., Bianchi, C., Bingham, R. G., Blankenship, D. D., Casassa, G., Catania, G., Callens, D., Conway, H., Cook, A. J., Corr, H. F., Damaske, D., Damm, V., Ferraccioli, F., Forsberg, R., Fujita, S., Gim, Y., Gogineni, P., Griggs, J. A., Hindmarsh, R. C., Holmlund, P., Holt, J. W., Jacobel, R. W., Jenkins, A., Jokat, W., Jordan, T., King, E. C., Kohler, J., Krabill, W., Riger-Kusk, M., Langley, K. A., Leitchenkov, G., Leuschen, C., Luyendyk, B. P., Matsuoka, K., Mouginot, J., Nitsche, F. O., Nogi, Y., Nost, O. A., Popov, S. V., Rignot, E., Rippin, D. M., Rivera, A., Roberts, J., Ross, N., Siegert, M. J., Smith, A. M., Steinhage, D., Studinger, M., Sun, B., Tinto, B. K., Welch, B. C., Wilson, D., Young, D. A., Xiangbin, C., and Zirizzotti, A. (2013). Bedmap2: Improved ice bed, surface and thickness datasets for Antarctica. *Cryosphere*, 7(1):375–393.

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, 102(477):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(3):502–523.

Golledge, N. R., Kowalewski, D. E., Naish, T. R., Levy, R. H., Fogwill, C. J., and Gasson, E. G. (2015). The multi-millennial Antarctic commitment to future sea-level rise. *Nature*, 526(7573):421–425.

Gomez, N., Mitrovica, J. X., Tamisiea, M. E., and Clark, P. U. (2010). A new projection of sea level change in response to collapse of marine sectors of the Antarctic Ice Sheet. *Geophysical Journal International*, 180(2):623–634.

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

Gramacy, R. B. and Lee, H. K. H. (2008). Bayesian treed gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103(483):1119–1130.

Greve, R. and Blatter, H. (2009). *Dynamics of ice sheets and glaciers*. Springer Science & Business Media.

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574.

Higdon, C. W., Mitra, R. D., and Johnson, S. L. (2013). Gene expression analysis of zebrafish melanocytes, iridophores, and retinal pigmented epithelium reveals indicators of biological function and developmental origin. *PLoS One*, 8(7):e67801.

Horton, B. P., Kopp, R. E., Garner, A. J., Carling, C., Khan, N. S., Roy, K., and Shaw, T. A. (2018). Mapping sea-level change in time, space and probability. *Annual Reviews of Energy and the Environment*, (July):1–41.

Hung Ying, Qian Peter Z.G., W. C. J. (2012). Statistical design and analysis methods for data center thermal management. In *Energy efficient thermal management of data centers(J. Yogendra and K. Pramod eds.)*, pages 497–513. Springer, New York.

IPCC (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*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Irvine, K. M., Gitelman, A. I., and Hoeting, J. A. (2007). Spatial designs and properties of spatial correlation: effects on covariance estimation. *Journal of agricultural, biological, and environmental statistics*, 12(4):450–469.

Joughin, I., Smith, B. E., and Medley, B. (2014). Marine ice sheet collapse potentially under way for the thwaites glacier basin, West Antarctica. *Science*, 344(6185):735–738.

Kaufman, C. G., Bingham, D., Habib, S., Heitmann, K., Frieman, J. A., et al. (2011). Efficient emulators of computer experiments using compactly supported correlation functions, with an application to cosmology. *The Annals of Applied Statistics*, 5(4):2470–2492.

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(484):1545–1555.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

Kopp, R. E., DeConto, R. M., Bader, D. A., Hay, C. C., Horton, R. M., Kulp, S., Oppenheimer, M., Pollard, D., and Strauss, B. H. (2017). Evolving Understanding of Antarctic Ice-Sheet Physics and Ambiguity in Probabilistic Sea-Level Projections. *Earth's Future*, 5(12):1217–1233.

Kukla, G. J., Bender, M. L., de Beaulieu, J. L., Bond, G., Broecker, W. S., Cleveringa, P., Gavin, J. E., Herbert, T. D., Imbrie, J., Jouzel, J., Keigwin, L. D., Knudsen, K. L., McManus, J. F., Merkt, J., Muhs, D. R., Müller, H., Poore, R. Z., Porter, S. C., Seret, G., Shackleton, N. J., Turner, C., Tzedakis, P. C., and Winograd, I. J. (2002). Last interglacial climates. *Quaternary Research*, 58(1):2–13.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics*, pages 386–404.

Levermann, A., Winkelmann, R., Nowicki, S., Fastook, J. L., Frieler, K., Greve, R., Hellmer, H. H., Martin, M. A., Meinshausen, M., Mengel, M., Payne, A. J., Pollard, D., Sato, T., Timmermann, R., Wang, W. L., and Bindschadler, R. A. (2014). Projecting Antarctic ice discharge using response functions from SeaRISE ice-sheet models. *Earth System Dynamics*, 5(2):271–293.

Li, R. and Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in gaussian kriging models. *Technometrics*, 47(2):111–120.

Liang, F., Cheng, Y., Song, Q., Park, J., and Yang, P. (2013). A resampling-based stochastic approximation method for analysis of large geostatistical data. *Journal of the American Statistical Association*, 108(501):325–339.

Little, C. M., Urban, N. M., and Oppenheimer, M. (2013). Probabilistic framework for assessing the ice sheet contribution to sea level change. *Proceedings of the National Academy of Science*, 110:3264–3269.

López, V. and Hamann, H. F. (2011). Heat transfer modeling in data centers. *International Journal of Heat and Mass Transfer*, 54(25-26):5306–5318.

Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71(1):135–146.

Marsh, L. C. and Cormier, D. R. (2001). *Spline regression models*, volume 137. Sage.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

Medley, B., McConnell, J. R., Neumann, T. A., Reijmer, C. H., Chellman, N., Sigl, M., and Kipfstuhl, S. (2018). Temperature and Snowfall in Western Queen Maud Land Increasing Faster Than Climate Model Projections. *Geophysical Research Letters*, 45(3):1472–1480.

Mengel, M., Levermann, A., Frieler, K., Robinson, A., Marzeion, B., and Winkelmann, R. (2016). Future sea level rise constrained by observations and long-term commitment. *Proceedings of the National Academy of Sciences*, 113(10):2597–2602.

Mercer, J. H. (1978). West antarctic ice sheet and co2 greenhouse effect: a threat of disaster. *Nature*, 271.

Mitrovica, J. X., Gomez, N., Morrow, E., Hay, C., Latychev, K., and Tamisiea, M. E. (2011). On the robustness of predictions of sea level fingerprints. *Geophysical Journal International*, 187(2):729–742.

Mitrovica, J. X., Hay, C. C., Kopp, R. E., Harig, C., and Latychev, K. (2018). Quantifying the sensitivity of sea level change in coastal localities to the geometry of polar ice mass flux. *Journal of Climate*, 31(9):3701–3709.

Nordman, D. J., Lahiri, S. N., and Fridley, B. L. (2007). Optimal block size for variance estimation by a spatial block bootstrap method. *Sankhyā: The Indian Journal of Statistics*, pages 468–493.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., and Sain, S. (2015). A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24(2):579–599.

Nychka, D., Wikle, C., and Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, 2(4):315–331.

Nychka, D. W. (2000). Spatial-process estimates as smoothers. *Smoothing and regression: approaches, computation, and application*, 329:393.

of PAGES, P. I. W. G. (2016). Interglacials of the last 800,000years. *Reviews of Geophysics*, 54:162–219.

Pagani, M., Liu, Z., LaRiviere, J., and Ravelo, A. C. (2009). High earth-system climate sensitivity determined from pliocene carbon dioxide concentrations. *Nature geoscience.*, 3(1):27,30.

Pal, J. S., Giorgi, F., Bi, X., Elguindi, N., Solmon, F., Gao, X., Rauscher, S. A., Francisco, R., Zakey, A., Winter, J., Ashfaq, M., Syed, F. S., Bell, J. L., Diffenbaugh, N. S., Karmacharya, J., Konar, A., Martinez, D., da Rocha, R. P., Sloan, L. C., and Steiner, A. L. (2007). Regional climate modeling for the developing world: The ictp regcm3 and regcnet. *Bulletin of the American Meteorological Society*, 88(9):1395–1410.

Pollard, D. and DeConto, R. M. (2009). Modelling West Antarctic ice sheet growth and collapse through the past five million years. *Nature*, 458(7236):329–332.

Pollard, D. and Deconto, R. M. (2012). Description of a hybrid ice sheet-shelf model, and application to Antarctica. *Geoscientific Model Development*, 5(5):1273–1295.

Pollard, D., DeConto, R. M., and Alley, R. B. (2015). Potential Antarctic Ice Sheet retreat driven by hydrofracturing and ice cliff failure. *Earth and Planetary Science Letters*, 412:112–121.

Pollard, D., Gomez, N., DeConto, R., and Han, H. K. (2018). Estimating Modern Elevations of Pliocene Shorelines Using a Coupled Ice Sheet-Earth-Sea Level Model. *Journal Geophysical Research: Earth Surface*, pages 1–13.

Pollard, D., Gomez, N., and Deconto, R. M. (2017). Variations of the Antarctic Ice Sheet in a Coupled Ice Sheet-Earth-Sea Level Model: Sensitivity to Viscoelastic Earth Properties. *Journal of Geophysical Research: Earth Surface*, 122(11):2124–2138.

Price, M., Handley, K., and Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education*, 36(8):879–896.

Rahmstorf, S. (2007). A semi-empirical approach to projecting future sea-level rise. *Science*, 315(5810):368–370.

Ritz, C., Edwards, T. L., Durand, G., Payne, A. J., Peyaud, V., and Hindmarsh, R. C. A. (2015). Potential sea-level rise from Antarctic ice-sheet instability constrained by observations. *Nature*, 528(7580):115–118.

Rougier, J. (2008). Efficient emulators for multivariate deterministic functions. *Journal of Computational and Graphical Statistics*, 17(4):827–843.

Rovere, A., Raymo, M., Mitrovica, J., Hearty, P., O'Leary, M., and Inglis, J. (2014). The mid-pliocene sea-level conundrum: Glacial isostasy, eustasy and dynamic topography. *Earth and Planetary Science Letters*, 387:27 – 33.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

RUE, H. and Tjelmeland, H. (2002). Fitting gaussian markov random fields to gaussian fields. *Scandinavian journal of Statistics*, 29(1):31–49.

Rutt, I. C., Hagdorn, M., Hulton, N., and Payne, A. (2009). The glimmer community ice sheet model. *Journal of Geophysical Research: Earth Surface*, 114(F2).

Santner, T. J., Williams, B. J., Notz, W. I., and Williams, B. J. (2003). *The design and analysis of computer experiments*, volume 1. Springer.

Schlegel, N.-j., Seroussi, H., Schodlok, M. P., Larour, E. Y., Boening, C., Limonadi, D., Watkins, M. M., Morlighem, M., and Broeke, M. R. V. D. (2018). Exploration of Antarctic Ice Sheet 100-year contribution to sea level rise and associated model uncertainties using the ISSM framework. *The Cryosphere Discussions*, (June):1–36.

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(3):743–758.

Schoof, C. (2007). Ice sheet grounding line dynamics: Steady states, stability, and hysteresis. *Journal of Geophysical Research: Earth Surface*, 112(3):1–19.

Sjöstedt-de Luna, S. and Young, A. (2003). The bootstrap and kriging prediction intervals. *Scandinavian Journal of Statistics*, 30(1):175–192.

Smola, A. J. and Bartlett, P. L. (2001). Sparse greedy gaussian process regression. In *Advances in neural information processing systems*, pages 619–625.

Snelson, E. and Ghahramani, Z. (2006). Sparse gaussian processes using pseudo-inputs. In *Advances in neural information processing systems*, pages 1257–1264.

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.

Stein, M. L. (2013). Statistical properties of covariance tapers. *Journal of Computational and Graphical Statistics*, 22(4):866–885.

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(2):275–296.

Strauss, B. H., Ziemlinski, R., Weiss, J. L., and Overpeck, J. T. (2012). Tidally adjusted estimates of topographic vulnerability to sea level rise and flooding for the contiguous united states. *Environmental Research Letters*, 7(1):014033.

Team, I. (2018). Mass balance of the Antarctic ice sheet from 1992 to 2017. *Nature*, 558:219–222.

Vaughan, D. G. (2008). West Antarctic Ice Sheet collapse - The fall and rise of a paradigm. *Climatic Change*, 91(1-2):65–79.

Vermeer, M. and Rahmstorf, S. (2009). Global sea level linked to global temperature. *Pnas*, 106(51):21527–21532.

Wang, Y. D. and Wu, C. J. (2017). Bayesian cubic spline in computer experiments. *Handbook of Uncertainty Quantification*, pages 477–495.

Weertman, J. (1974). Stability of the junction of an ice sheet and an ice shelf. *Journal of Glaciology*, 13(67):311.

Wegman, E. J. and Wright, I. W. (1983). Splines in statistics. *Journal of the American Statistical Association*, 78(382):351–365.

Wikle, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of spatial statistics*, pages 114–125. CRC Press.

Wold, S. (1974). Spline functions in data analysis. *Technometrics*, 16(1):1–11.

Ying, Z. et al. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *The Annals of Statistics*, 21(3):1567–1590.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.

Zhang, H. and Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, 92(4):921–936.

Zhao, Y., Amemiya, Y., and Hung, Y. (2018). Efficient gaussian process modeling using experimental design-based subagging. *Statistica Sinica*, 28:1459–1479.

Zhu, Z. and Stein, M. L. (2006). Spatial sampling design for prediction with estimated parameters. *Journal of agricultural, biological, and environmental statistics*, 11(1):24.

Zickfeld, K., Solomon, S., and Gilford, D. M. (2017). Centuries of thermal sea-level rise due to anthropogenic emissions of short-lived greenhouse gases. *Proceedings of the National Academy of Sciences*.