Title
Hierarchical Bayesian Spatial Modeling for Quantifying Climate Uncertainty

Author
Baugh, Samuel J

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Hierarchical Bayesian Spatial Modeling for Quantifying Climate Uncertainty

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Statistics

by

Samuel Baugh

2022

ABSTRACT OF THE DISSERTATION

Hierarchical Bayesian Spatial Modeling for Quantifying Climate Uncertainty

by

Samuel Baugh
Doctor of Philosophy in Statistics
University of California, Los Angeles, 2022
Professor Karen A. McKinnon, Chair

In the age of global warming, there is a crucial need to accurately assess uncertainty levels when analyzing observed changes in the climate. For many climate problems, the development of statistical methods that appropriately account for uncertainty is challenging due to the complexity of the underlying climate processes and the various sources of uncertainty involved. This thesis addresses methodological challenges in modeling uncertainty for two climate problems with important real-world applications. The first problem is concerned with quantifying the heat content of the global ocean and its change over time. Understanding the trend in ocean heat content is particularly important as it informs estimates of transient climate sensitivity, a physical parameter that largely determines the amount of warming that will be expected in the years to come. This problem is nevertheless made difficult by the challenge of representing the complex covariance structure of the ocean heat content field, as well as the challenge of quantifying the uncertainty in the estimation of this structure. The second problem is concerned with separating the influence of warming caused by human activities from natural variability in the observed climate, a problem that is often referred to as climate change "detection and attribution". While various sources of uncer-

tainty in this problem have been addressed in the literature, recent results have suggested that commonly-used methods under-estimate uncertainty in their conclusions. Producing reliable detection and attribution confidence intervals is difficult in part due to the challenge of modeling the uncertainty in the estimation of the natural variability covariance structure from limited climate model simulations.

This thesis proposes methods for addressing statistical challenges in these two problems with respect to three overarching themes. The first theme is the use of spatially-coherent statistical models to represent the covariance structures of the underlying physical processes. For the ocean heat content problem, a novel cylindrical kernel-convolution Gaussian process model is developed to flexibly represent the complex spatial correlation patterns of the global ocean heat content field. For the detection and attribution problem, a Laplacian basis vector parameterization of the covariance matrix is proposed to enforce spatially-coherent correlation patterns. This parameterization is also able to avoid the uncertainty in the traditional approach of estimating principal component vectors from limited numbers of climate model runs. The second theme is the use of hierarchical Bayesian models to propagate the uncertainty in estimating the covariance structure to the final results. In the ocean heat content problem, the spatially-varying parameter fields describing the kernel-convolution Gaussian process are themselves modeled as Gaussian processes in a hierarchical framework. This allows for the uncertainty in estimating these parameters to be propagated to the final posterior distribution for the ocean heat content trend. In the detection and attribution problem, the parameters of the Laplacian parameterization of the covariance matrix, as well as the number of Laplacians to use, are both represented in a Bayesian hierarchical framework that prioritizes the accurate modeling of uncertainty. Finally, the third theme concerns the evaluation of the statistical properties of the Bayesian posterior distributions. For the ocean heat content problem, this is done using cross-validation on the observations with respect to a metric for evaluating both the mean and uncertainty implied by the posterior predictive distributions. For the detection and attribution problem, climate model simulations are

used to evaluate the accuracy of the posterior means and credible intervals produced by the proposed methods in the context where the true value can be assumed to be known.

Chapter 1 begins by introducing the broader context and implications of the two climate problems and proceeds to give a brief overview of the three statistical themes. Chapter 2 develops the proposed methodology for the ocean heat content problem in a restricted context focusing on spatial variability. A cross-validation study is presented showing that the proposed framework achieves higher accuracy in the predictive posterior distributions than a commonly-used previous method as well as simpler Bayesian approaches. This framework is then extended to the full spatio-temporal context in Chapter 3 and is applied to the quantification of the trend in ocean heat content from 2007 to 2021. The detection and attribution problem is addressed in Chapter 4, where a climate model validation study shows that the proposed approach achieves higher accuracy in the posterior mean and more accurate credible intervals than a traditional approach. While the validation results for each of these proposed methods show quantitative improvements over previous approaches, the results suggest several promising opportunities for additional improvements and extensions. Several of these potential avenues for future research are discussed in Chapter 5.

The dissertation of Samuel Baugh is approved.

Marilyn Raphael

Frederick Schoenberg

Mark Handcock

Karen A. McKinnon, Committee Chair

University of California, Los Angeles

2022

To my mom and dad, who have supported and encouraged me throughout the way.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Over the 2021-2022 school year I was supported by the UCLA Dissertation Year Fellowship.

Chapter 2 is adapted from the following article to appear in *Annals of Applied Statistics*:

Samuel Baugh and Karen McKinnon. Hierarchical Bayesian Modeling of Ocean Heat Content and Its Uncertainty. *Annals of Applied Statistics.* 2022.

This manuscript was written by Samuel Baugh under the advisement of Karen McKinnon.

First and foremost I would like to thank my advisor Karen McKinnon, without whom this thesis would not have been possible, for her guidance and encouragement throughout this process and for helping me contribute to the exciting and growing world of climate statistics. Thanks for always giving me prompt and insightful feedback on my work, for encouragement throughout the ups and downs of the research process, and for creating a great group for statistical climate research group at UCLA.

I would like to thank Mark Handcock for his statistical insights and for his unwavering support for my research and career goals. I would also like to thank Rick Schoenberg and Marilyn Raphael for their guidance as members of my committee.

I would like to thank Minjeon Jeong for providing me with opportunities to develop my knowledge of Bayesian analysis on problems outside of my main application areas.

I would like to thank Professors Mahtash Esfandiari, Nicolas Christou, and Rob Gould for showing me that teaching is a core part of what it means to be a good statistician. I would also like to thank Laurie Leyden for her incessant work and dedication to keeping the stats department running effectively.

I would like to thank fellow Ph.D. students Sydney Kahmann, Andy Kaplan, Ami Wulf, Chris Strohmeier, and Stephanie Stacey for their companionship both inside and outside statistics, as well as fellow TACs Lenny Wainstein and Stephen Smith, fellow consultant

Zack Stokes, fellow stats GSA members Harrison Katz, Carlos Cinelli, Thomas Maierhofer and Sasha Farzin-Nia, and my math camp co-instructors Zhanhao Peng and Jiayi Li.

I would like to thank Wenwen Kong for doing most of the work in organizing the stats-climate journal club, I have found this group to be an invaluable opportunity to learn about and discuss exciting research and was the source of the idea for the second part of my dissertation. I would also like to thank my other lab group members Kyle McEvoy, Suqin Duan, Kim Wang, and Nathan Hwangbo for providing a great forum for discussing statistical climate research and for providing valuable feedback on my practice presentations.

I would like to thank my roommate Alec Snyder for always patiently listening to my statistical ramblings and for introducing me to some of the great food available in Los Angeles.

I would like to thank my Master's thesis advisor Michael Stein, for encouraging me to pursue my PhD at UCLA and for helping me begin my passion for spatial and environmental statistics. I would like to thank Mikkael Kuusela for providing the inspiration for my approach in the first section of my dissertation and for organizing the statistical oceanography discussion group. I would also like my fellow UChicago BS-MS student Peter Gao for organizing the virtual spatio-temporal statistics reading group over the pandemic.

I would like to thank Mark Risser for his advice and encouragement on the future continuation of my work on statistical problems in climate change detection and attribution.

Finally I would also like to thank my mom and dad for their unwavering support, for giving me a place to shelter during the pandemic, and for reminding me that "the best dissertation is a done dissertation". I would also like to thank my siblings Matthew, Josh, Jinny, and Amy for their support, my niece Morgan for being a dependable source of adorable pictures in the family groupchat, and to our cat Cinnamon for being my reliable research assistant during the pandemic.

VITA

2013-2017 **B.S.** Mathematics, University of Chicago.

2016-2017 **M.S.** Statistics, University of Chicago.

2017-2021 **Teaching Assistant**, UCLA Department of Statistics

2019-2022 **Graduate Research Assistant**, UCLA Graduate School of Education and Information Sciences

2019-2022 **Graduate Student Consultant**, UCLA Statistical Consulting Center

PUBLICATIONS

Samuel Baugh and Karen McKinnon. Hierarchical Bayesian Modeling of Ocean Heat Content and Its Uncertainty. *Annals of Applied Statistics.* (2022).

Minjeong Jeon, Ick Hoon Jin, Michael Schweinberger, and **Samuel Baugh**. Mapping unobserved item- respondent interactions: A latent space item response model with interaction map. Psychometrika, 86, pp.378–403 (2021).

Samuel Baugh and Michael L. Stein. Computationally efficient spatial modeling using recursive skele- tonization factorizations. Spatial Statistics, 27, pp.18-30 (2018).

Samuel Baugh, Gruia Calinescu, David Rincon-Cruz, and Kan Qiao. Improved Algorithms

for Two Energy-Optimal Routing Problems in Ad-Hoc Wireless Networks. Proceedings of IEEE International Conferences on Big Data and Cloud Computing, pp.509-516 (2016).

Kelsey Roelofs, Valerie Juniat, Luke Ledbetter, Michael O'Rourke, Daniel Helmy, Thomas Hardy, Jean Lee, Mahtash Esfandiari, **Samuel Baugh**, Dinesh Selva, and Daniel Rootman. Improving the accuracy of pre-operative diagnosis in well-circumscribed orbital tumors: A multi-center study. Ophthalmic Plastic and Reconstructive Surgery (2021).

CHAPTER 1

Introduction

While there is a long-established scientific consensus that increasingly elevated greenhouse gas concentrations from human activities have and will continue to warm the Earth's climate system [Masson-Delmotte et al., 2021], there is significant uncertainty in understanding exactly how human activities have influenced the observed climate state and what this implies for future warming. Two examples of climate problems where uncertainty imposes major challenges are the estimation of equilibrium climate sensitivity and the "detection and attribution" of the climate change signal. Equilibrium climate sensitivity is defined as the expected long-term increase in global temperatures in response to a doubling of atmospheric carbon dioxide levels [Knutti et al., 2017]. Since the first approaches in the late 1800s [Arrhenius, 1896], the estimation of this parameter has carried notoriously high levels of uncertainty, with the current best estimate range of 1.5 °C to 4.5 °C being essentially unchanged since the 1970s [Charney et al., 1979]. Transient climate response, which is the parameter analogous to equilibrium climate sensitivity at the more policy-relevant scale of decades, is slightly more constrained, with a 1 °C to 2.5 °C best-estimate range. Regardless, the ends of these ranges have vastly different implications for the environment and society [Adger et al., 2007], and as such there is a crucial need to further incorporate empirical observations in estimation and uncertainty quantification [Caldwell et al., 2016, Tokarska et al., 2020]. As 90% of the energy imbalance caused by greenhouse gases is expected to be stored in the world's oceans [Trenberth et al., 2014], quantifying how much heat the oceans have gained in the observational record plays an important role in quantifying equilibrium and transient climate sensitivities [Knutti and Tomassini, 2008, Meyssignac et al., 2019]. How-

ever, historically limited observations and the complexity of the ocean temperature field have rendered the estimation of ocean heat uptake challenging [Cheng et al., 2017a, Resplandy et al., 2019].

Another important problem where uncertainty plays a key role is in climate change detection and attribution, the goal of which is to detect the influence of greenhouse gas emissions in the observed climate and to determine the extent to which human influences, as opposed to natural variability, are responsible for observed changes. Optimal fingerprinting techniques within the detection and attribution literature have played a crucial role in establishing the link between greenhouse gas concentrations and observed changes in global temperatures [Hegerl et al., 1994, Allen and Tett, 1999]. The concepts of variability and uncertainty are paramount in optimal fingerprinting methodology, which must take into account uncertainty induced by the natural properties of the climate system itself, from the estimation of the observed climate state from limited observations, and from estimating key parameters, including the structure of variability, from climate models.

Iterative improvements have been made over the last two decades to incorporate the various sources of uncertainty in optimal fingerprinting methodology [Hammerling et al., 2019]. However, recent research has suggested that established methods under-estimate the true uncertainty in the problem and as such produce potentially misleading conclusions [Li et al., 2021, McKitrick, 2022]. Since establishing the responsibility of human activities for climate impacts has important consequences for mitigation and adaptation efforts, it is crucial that conclusions are made with reliable uncertainty estimates. While the results of optimal fingerprinting studies do not play a decisive role in the scientific consensus on global warming, which is supported by multiple lines of empirical and theoretical evidence, links between human activities and climate changes at smaller scales, other climate variables such as precipitation, and climate extremes are less statistically and scientifically obvious [Christidis et al., 2012, Knutson and Zeng, 2018, Wang et al., 2021]. As policymakers and the public at large are particularly interested in climate change impacts at local scales,

the development of accurate and reliable optimal fingerprinting frameworks has significant real-world consequences.

These two climate problems are important for understanding the anticipated magnitude of future climate change and the extent to which human activities are responsible for these changes. However, in both problems, high levels of uncertainty pose a major challenge in quantifying the evidence from available information. Across these problems, uncertainty can be thought of as primarily arising from three broad sources. The first is the inherent variability of the natural climate system. While the global climate system is driven by physical dynamics, the sensitivity of the equations describing these dynamics to small changes in initial conditions renders the global ocean and atmospheric systems effectively probabilistic [Slingo and Palmer, 2011]. The second source of uncertainty is caused by the sampling of finite numbers of observations from continuous processes. The third source of uncertainty arises when output from climate model simulations is used to incorporate physical knowledge into the analysis. Climate models, which produce simulations of key climate variables using numerical solutions to the underlying dynamical equations, provide an invaluable source of information for understanding the climate system. These resources are particularly crucial to the detection and attribution problem given that the forced signal is confounded with unforced variability in the observations. However, the use of these simulations introduces uncertainty through errors caused by the numerical approximation of certain climate processes, in addition to uncertainty from the finite amount of simulation output [Schneider and Dickinson, 1974, Washington and Parkinson, 2005, Griffies, 2018].

Statistics, as the science of using the principles of probability and randomness to quantify uncertainty, has the ability to play a fundamental role in understanding these and other important climate questions. Motivated at a high level by the two important applications introduced above, this thesis will investigate issues in the development of statistical techniques to model variability and uncertainty in climate science. Improvements to existing approaches will be developed along three themes. The first theme concerns the development

of spatially-correlated covariance models on the sphere to represent variability in physical processes. The second theme is on the development of hierarchical Bayesian frameworks to integrate multiple sources of uncertainty into a posterior distribution. Finally, the third theme regards the use of model validation techniques to evaluate both the mean and the uncertainty in the posterior distributions produced by the statistical methods.

The first theme explored in this thesis regards spatial modeling of the covariance structure of climate processes. In the spatial statistics literature, Gaussian processes are a well-established way to assign a probability distribution to continuous fields which vary at least somewhat smoothly over space [Krige, 1951, Matheron, 1963, Cressie, 1990, Banerjee et al., 2003, Stein, 1999a, Cressie, 2015]. Despite the attractive theoretical properties of Gaussian process models, and their suitability for modeling diverse climate phenomena, they have been very infrequently employed in the climate science literature. This reticence is primarily due to historical difficulties in representing complex processes on the spherical domain as well as the oftentimes prohibitive computational load encountered when applying Gaussian processes to reasonably sized climate datasets. The development of kernel-convolution Gaussian processes has allowed for flexible modeling of complex spatially-varying covariance structures as are often found in climate processes [Higdon, 1998a, Paciorek and Schervish, 2006a, Risser, 2016]. However, such methods are most commonly applied to the Euclidean domain and encounter additional challenges in computation and interpretation when extended to the sphere [Jeong and Jun, 2015, Li and Zhu, 2016a, Guinness and Fuentes, 2016]. On the computational side, recent developments in approximating the Gaussian process likelihoods have allowed for a viable way to model datasets larger than previously possible. In particular, Vecchia processes, originally proposed by Vecchia [1988], approximate the full Gaussian process by restricting conditional dependence relations in the data to subsets of nearest neighbors. This allows for computation time to be balanced with accuracy by tuning the maximum size of the conditioning sets. Recent theoretical and methodological advancements have allowed for Vecchia processes to fully substitute Gaussian processes in many

circumstances [Katzfuss et al., 2020a, Guinness, 2021, Katzfuss and Guinness, 2021, Zilber and Katzfuss, 2021, Jimenez and Katzfuss, 2022]. Recursive skeletonization factorizations, which compress numerically-low rank blocks of the Gaussian process covariance matrix, are another computational technique for the fast computation of Gaussian process likelihoods with a trade-off between computation time and accuracy [Minden et al., 2017, Baugh and Stein, 2018]. Despite these developments, Gaussian process models have the general computational downside in that the entire covariance matrix approximation must be re-computed for every configuration of the parameters. An alternative technique for representing spatially-coherent patterns in a covariance structure would be to use fixed basis functions. When the domain is Euclidean, the Fourier transform can be used to obtain a basis function representation of a stationary Gaussian process [Paciorek, 2007]. For the cylindrical domain, an analogous approach would be to use spherical harmonics, which can be obtained from the eigenfunctions of the Laplace operator on the sphere [Courant and Hilbert, 2008, DelSole and Tippett, 2015].

The second theme is the use of hierarchical Bayesian frameworks to propagate multiple sources of information and uncertainty to the final inference. It should be noted that while many statisticians and scientists have advocated for the use of Bayesian methods in climate problems [Berliner et al., 2000, Smith et al., 2009, Mann et al., 2017], historically most statistical methods employed in climate science have fallen under a frequentist philosophy [Shepherd, 2021]. This presumes that parameters of interest have true, fixed values that are unknown to the observer. Statistical methods are then used to quantify the uncertainty which results from sampling from a distribution with unknown parameters. Generally, to obtain distributional forms for the estimators, a null hypothesis is assumed, and if the stochastic estimators have a low probability conditional on this hypothesis then it is rejected in favor of a conversely defined alternative. This philosophy, while appropriate for scenarios such as population inference where a "true" value is known to exist, is less justified when the underlying quantities of interest are better thought of as probabilistic. Arguably, most

quantities in climate science can be represented as random variables, either by their physical nature or by the fact that knowledge of their value can be described with an uncertainty distribution. Under this point of view, Bayesian inference appears more appropriate for the analysis of climate problems as Bayesian methodology views all unknown quantities as random variables. Using Bayesian theory and established inference techniques, distributions can be updated conditional on different sources of variability and uncertainty into a posterior distribution for the quantity of interest. Furthermore, since Bayesian frameworks allow for the distributions of the parameters of a statistical model to be treated as random variables, the uncertainty in their estimation can be propagated to the final result. As such, Bayesian frameworks are a natural fit for the analysis of climate processes, where multiple sources of uncertainty are present and nearly all components can be viewed as coming from probabilistic distributions.

A major challenge in applying Bayesian methods is the number of methodological choices to be made in constructing an appropriate model [Gelman and Hill, 2006]. The first is choosing the likelihood distribution of the data. In the common case where the distributional form is taken to be normal, this requires specifying forms for the mean and variance of the process. The second choice is the prior distribution. The prior distribution can either be "informative", in explicitly containing prior information about the quantities under consideration, or uninformative in not explicitly containing prior information, although from a technical perspective no prior is truly uninformative [Van Dongen, 2006]. A third set of choices to be made in conjunction with the above two items is in the development of a hierarchical structure for the components of the statistical model. The simplest Bayesian hierarchical models generally assign priors to the hyper-parameters describing the prior distributions of the parameters, which themselves determine the likelihood of the data. This allows for the distributions of the parameters themselves to have properties informed by the data. Different levels of a Bayesian hierarchical framework can also allow for choices in the likelihood function to be explicitly quantified while allowing for uncertainty from different sources to

be incorporated in the inference framework. With appropriate methodological choices, hierarchical Bayesian models provide a powerful technique to quantify uncertainty in climate problems.

Due to the diverse sources of uncertainty present in climate problems, it is particularly challenging to assess whether uncertainty estimates correctly represent the true variability present. This motivates the third theme regarding the use of statistical model validation techniques to evaluate both the center and the uncertainty levels of posterior distributions. The continuous-ranked probability scores metric, or CRPS [Matheson and Winkler, 1976], was popularized within the climate community to evaluate the accuracy of probabilistic weather predictions [Hersbach, 2000]. In order to apply this metric, the "true" value of the quantity being estimated or predicted must be known. In the statistics and climate literature, a common way to evaluate predictions against true values is through cross-validation, where certain segments of the data are withheld and the model conditioned on the remaining data is used to predict the withheld value or values [Stone, 1978, Bates et al., 2021]. In certain contexts, climate model simulation data can be used as a substitute for the observations in order to evaluate the properties of the statistical model with a known truth determined by the climate model. Recently there have been increases in the availability and size of initial-condition ensembles, which are collections of output from a fixed climate model configuration run under various perturbations of the initial conditions [Kay et al., 2015, Sriver et al., 2015, O'Neill et al., 2016, Stein, 2020]. These ensembles provide a valuable source of information for evaluating the properties of statistical methods in view of the "true" variability implied by the climate models.

These three themes will be explored through the development of statistical methodology for two climate science problems motivated by the broader questions introduced at the beginning of this section. The first problem concerns estimating the rate of ocean heat uptake over the past several years. Because ocean heat uptake is a crucial quantity for understanding equilibrium and transient climate sensitivity, it is important to not just achieve a reliable

estimate of this quantity but to accurately estimate the uncertainty in its estimation. Recent increases in data made available through the Argo program [Riser et al., 2016] have allowed for an unprecedented view into ocean heat content, however, the complexity of the covariance structure of the ocean heat content field makes statistical modeling challenging. A statistical approach needs to account for three sources of uncertainty; the natural variability in the field itself, the sampling variability induced from finite observations of the underlying process, and the variability induced from having to estimate the parameters describing the covariance structure of the process.

In Chapter 2, a Bayesian hierarchical Gaussian process model for quantifying the uncertainty in the ocean heat content field and its trend over time is developed. At the first level of the hierarchical structure, the heat content field is assumed to be distributed according to a spatial Gaussian process. To overcome difficulties in modeling the complex spatially-varying properties of the ocean heat content covariance structure, a key innovation is the use of a cylindrical kernel-convolution construction for the Gaussian process which allows for the correlation parameters, namely the process variance, nugget variance, and latitudinal and longitudinal correlation length scales, to vary over space. On the second level of the hierarchical model, the parameters that describe the correlation fields, as well as the mean and trend fields, are themselves assumed to follow spatial Gaussian processes. The behavior of these prior distributions is determined by hyper-priors whose values are informed by the data itself. This Gaussian process model is able to account for the process variability in the ocean heat content field through the estimation of the spatially-varying covariance structure. Furthermore, the spatial nature of this process allows for the interpolation of the heat content field to unobserved values using the theory of kriging. Finally, the representation of the parameter fields as random processes themselves in the second level of the hierarchy allows for the uncertainty induced from estimating the complex correlation structure to be explicitly incorporated in the final posterior distribution. These properties are demonstrated through the application in Chapter 2 to the trend in ocean heat content as gauged by data

restricted to the month of January, where the data restriction allows for the model to focus on the spatial aspects of the correlation structure and mean-field. This framework is extended in Chapter 3 to a spatio-temporal model with a seasonally and spatially varying mean-field applied to data from all months over the period 2007 to 2021.

The second problem addressed in this thesis is modeling the covariance matrix for the detection and attribution problem. Similar to the statistical modeling of ocean heat content, this problem requires estimating the true state of the observed climate from limited observations. In contrast, the detection and attribution problem requires an additional source of information, specifically the expected true effect of different forcing patterns. Here, "forcing" is defined as any change which could impact the climate, with the forcing of increasing greenhouse gas concentrations of primary concern. Since the goal is to distinguish between changes in the climate caused by the forcing and changes that are caused by natural variability, there is an imperative need to obtain a physically-informed estimate of the natural variability structure. Physics-based climate models are used in the literature to estimate the forced signal and the natural variability structure, generally through "historical" or forced runs and "pre-industrial" or unforced runs respectively. While climate models have provided an invaluable source of information for assessing the detection and attribution of climate change, their inclusion presents another source of uncertainty that needs to be taken into account within the statistical treatment. These sources of uncertainty include both the "sampling" uncertainty from having only a finite number of model runs to estimate the desired features of the climate field, and from differences between the climate models and the true climate system.

Chapter 4 develops a Bayesian hierarchical model for the detection and attribution of observed climate changes that can take into account the uncertainty from estimating the structure of internal variability from limited climate model simulations. The first level of the hierarchy represents the observed state of the climate as the estimated forced signal plus a residual term following the distribution of unforced natural variability. In the second

level of the hierarchy, the natural variability covariance matrix is modeled as the distribution of pre-industrial simulation data. To obtain a distributional form for the internal variability structure that produces reliable detection and attribution inferences, this chapter introduces two innovations to previous Bayesian approaches. The first is to parameterize the covariance matrix with Laplacian basis functions in order to enforce spatially-coherent correlation patterns. This provides an advantage over traditional methods, which use principle components to parameterize the covariance matrix, since such components are calculated conditional on a set of climate model runs, and as such the uncertainty in their estimation is not easily incorporated into the inference procedure. The second innovation is the use of a χ^2 re-parameterization of the likelihood function for modeling the number of Laplacian basis vectors. This allows for the selection of the number of covariance parameters, which is traditionally done outside of the inference procedure, to be expressed through the top level of the Bayesian hierarchy. An extensive climate model validation study shows that this method yields more reliable posterior distributions for the detection and attribution parameter than other approaches.

The three statistical themes, of spatial covariance modeling, hierarchical Bayesian modeling, and Bayesian model validation are interwoven throughout these two applied problems. The solutions to both problems involve parameterizations of the covariance structure that enforce spatial correlation. For the ocean heat content problem this is done through a Gaussian process model, and for the detection and attribution problem this is done through a Laplacian basis function parameterization. Both problems then propagate the uncertainty in estimating the covariance structure to the final results with a hierarchical Bayesian framework. While in the ocean heat content problem uncertainty in estimating the covariance structure comes from the observations alone, the approach for the detection and attribution problem also incorporates uncertainty from the use of climate model simulations. Finally, the third theme is expressed through cross-validation in the ocean heat content problem and through climate model validation in the detection and attribution problem. Results

using these validation techniques show that the spatial covariance models and the Bayesian hierarchical frameworks developed for these problems yield quantitative improvements over previous approaches.

The rest of this thesis will be organized as follows. The spatial-only model for quantifying ocean heat content will be developed in Chapter 2. This chapter is largely taken from the paper [Baugh and McKinnon, 2022], which has been accepted for publication by *The Annals of Applied Statistics*. Chapter 3 then extends the spatial model to the spatio-temporal setting and applies the framework to all data between 2007 and 2021. Then, the detection and attribution framework is developed and applied to changes in near-surface air temperatures from 1990 to 2015 in Chapter 4. The thesis concludes with a discussion of directions for future research in Chapter 5.

CHAPTER 2

Hierarchical Bayesian Modeling of Ocean Heat Content

The accurate quantification of changes in the heat content of the world's oceans is crucial for understanding the effects of increasing greenhouse gas concentrations. The Argo program, consisting of Lagrangian floats that measure vertical temperature profiles throughout the global ocean, has provided a wealth of data from which to estimate ocean heat content. However, creating a globally consistent statistical model for ocean heat content remains challenging due to the need for a globally valid covariance model that can capture complex nonstationarity. This chapter develops a hierarchical Bayesian Gaussian process model using kernel convolutions with cylindrical distances to allow for spatial non-stationarity in all model parameters while using a Vecchia process to remain computationally feasible for large spatial datasets. This approach can produce valid credible intervals for globally integrated quantities that would not be possible using previous approaches. These advantages are demonstrated through the application of the model to Argo data, yielding credible intervals for the spatially-varying trend in ocean heat content that accounts for both the uncertainty induced from interpolation and from estimating the mean-field and other parameters. The proposed model out-performs an out-of-the-box approach as well as other simpler models in a cross-validation study using continuous ranked probability scores (CRPS) to evaluate both the mean and uncertainty values of the predictive posterior distributions. The code for performing this analysis has been developed into an R package **BayesianOHC**.

2.1 Introduction

As 90% of the energy imbalance caused by increased greenhouse gas concentrations is expected to be stored in the world's oceans [Trenberth et al., 2014], measuring changes in ocean heat content is essential to the scientific understanding of climate change. However, obtaining accurate estimates of ocean heat content has historically been difficult due to a lack of data. To help address this problem, since 2002 the Argo program has been coordinating and maintaining a network of thousands of Lagrangian floats that passively move throughout the ocean measuring temperature and salinity [Riser et al., 2016]. While Argo floats achieve dense sampling in the vertical dimension, quantifying ocean heat content remains challenging since the number of Argo observations is small in comparison to the vastness of the global ocean. As estimates of ocean heat content inform scientific knowledge about other climate processes, it is particularly important to not just obtain an estimate of the ocean heat content trend but to assess the uncertainty in its estimation.

Several techniques have been used in the climate science literature to quantify ocean heat content using Argo data. One group of techniques is referred to as simple-gridding [Meyssignac et al., 2019] and includes approaches such as von Schuckmann and Le Traon [2011] and Gouretski [2018]. These methods compute ocean heat content at each point of a grid over the ocean's surface by taking the average value of the observations within the corresponding grid-box. For grid-boxes with few observations, the heat content value is assumed to either have an anomaly value of zero or be equal to the global average value of all observations. Standard errors are obtained by computing the empirical standard deviation of the observations in each grid-box; for grid-boxes with few observations, the global standard deviation over all observations is used. A major drawback of this method is that it does not take advantage of the correlation structure in the data, which will lead to more variable estimates of the mean-field and precludes a globally consistent estimate of the uncertainty in heat content.

Another group of approaches, including Levitus et al. [2012] and Ishii et al. [2017], estimate the heat content anomaly at a particular point using values observed at nearby locations. Levitus et al. [2012] estimate heat content with a weighted sum of observations within a fixed radius, where the weights decrease with the distance between the observation locations at a rate determined by a fixed correlation length scale. Ishii et al. [2017] estimate unobserved values by minimizing a cost function that accounts for the influence of nearby observations by using an inverse-distance weighting also with a fixed correlation length scale. A major drawback of these techniques is that the correlation length scales are not informed by the data and must be chosen in advance of the interpolation. Standard errors are estimated in these techniques by computing empirical variances and covariances on the observations partitioned into grid-boxes. While this standard error calculation takes into account the fact that observations in nearby grid-boxes will be correlated, the standard errors do not take into account the spatial distribution of observations within each grid-box. Furthermore, like the simple gridding approach, this method does not produce a consistent covariance estimate across the global domain.

Other approaches use data from climate models to aid in the estimation of ocean heat content. The approach of Cheng and Zhu [2016] interpolates the data using covariances obtained from output of the Coupled Model Intercomparison Project Phase 5. Approaches such as Balmaseda et al. [2013] interpolate the heat content field using climate model runs constrained by the observed values. Both of these methods suffer from the drawback that they are subject to the effect of climate model error on the covariance structure which can be difficult to assess and correct for [Palmer et al., 2017].

Gaussian processes are a powerful technique for expressing a spatial field with a probability model. The key assumption of the Gaussian process is that every linear combination of values in a spatial field follows a multivariate normal distribution. Then the correlation between each pair of observations is determined by their locations as well as by parameters that can be estimated from the data. Under the Gaussian process assumption, the value at

an unobserved location conditioned on a set of observed values has a normal distribution whose mean is the linear combination of the observations [Stein, 1999b]. The uncertainty in estimating the value at the unobserved location can then be obtained through the standard deviation of the conditional normal distribution.

The Gaussian process interpolation is similar to the weighted sum method used in Levitus et al. [2012] in that the estimates are given by a linear combination of the observations. However, the Gaussian process imposes a probability model on the spatial field, from which it can be shown mathematically that the conditional mean at an unobserved location is the best linear unbiased predictor [Stein, 1999b]. Another advantage of assigning a probability model to the spatial field is that it allows for the correlation parameters to be determined through statistical inference, rather than chosen in an ad-hoc fashion. The uncertainty in the estimation of the parameters can also be assessed using statistical inference techniques.

Recent work in the statistics literature by Kuusela and Stein [2018] uses Gaussian processes on the Argo data to assess variability in the temperature fields at fixed depths. One challenge of applying Gaussian processes is that they were originally developed for fields whose covariance parameters are constant over the domain. However, the statistical properties of the ocean temperature field are known to vary substantially between different regions of the global domain [Cheng et al., 2017b]. Kuusela and Stein [2018] account for non-stationarity by assuming local stationarity within windows of a fixed size centered around each point of a grid partitioning the ocean. This allows the model to be fit independently at each grid-point yielding an estimate and uncertainty range for the value at the center of each window. However, as ocean heat content is a globally integrated quantity, assessing its uncertainty cannot be accomplished with this local method. Instead, a covariance model that is valid over the entire domain is required, which is the focus of this work.

It is easier to model ocean heat content, defined as the integral of the three-dimensional temperature field at a particular time [Dijkstra, 2008], than modeling the temperature field directly. Thus, the problem is reduced to fitting a non-stationary model to the surface of

the sphere. Even so, modeling non-stationary fields on a sphere is a major theoretical and computational challenge [Jeong et al., 2017]. Kernel-convolution methods introduced by Higdon [1998b] and further developed by Paciorek and Schervish [2006b] and Risser [2016] provide a flexible way to construct a Gaussian process with parameters that vary over the domain. In these methods, each location in the field is assigned a parameterized kernel function and the global model is obtained by computing convolutions between the kernels at any two points. When the kernel parameters vary smoothly, they describe the local behavior of the process at that point, yielding a clear interpretation of their meaning.

While the convolutions have easily expressed forms when the domain is assumed to be Euclidean, they are more challenging to compute when the domain is spherical. Computing convolutions using numerical approximation of the integrals as in Jeong and Jun [2015] and Li and Zhu [2016b] is computationally intensive and even with pre-computation is infeasible on the thousands of observations in the Argo dataset. Various other non-stationary correlation functions have been developed for Gaussian processes on the sphere, such as deformations of the sphere introduced by Sampson and Guttorp [1992], spherical harmonics used by Hitczenko and Stein [2012], and stochastic partial differential equations used by Lindgren et al. [2011], however, the parameters of these covariance functions do not have the intuitive physical interpretation afforded by kernel convolution methods. Recently the **BayesNSGP** package (Turek and Risser [2019], Risser and Turek [2020], and Risser [2020]) offers the ability to use kernel convolutions in a Bayesian framework to model highly non-stationary processes in a computationally efficient manner. This package allows for non-stationary anisotropic modeling on the two-dimensional Euclidean domain and non-stationary isotropic modeling on the spherical domain.

The rest of this chapter develops a kernel-convolution spatial Gaussian process treating the surface of the ocean as a cylinder to produce a model that has spatially-varying and interpretable parameters while remaining computationally feasible. The spatially-varying parameter fields are represented at the second level of the hierarchical model as Gaussian

processes themselves with hyper-parameters informed by the data. The Gaussian process prior for the parameters allows for the longitudinal correlation length scale to vary with respect to the changing radius of the Earth with latitude; this, in combination with the fact that there are no Argo observations close to the poles, allows for the use of the cylindrical distance rather than more computationally challenging spherical distance metric. The methodology used in this chapter, including the covariance model, the Bayesian hierarchical framework, and the MCMC model-fitting procedure is described in detail in Section 2.3. Fitting this highly flexible Bayesian model on the number of points in the Argo dataset is still a computational challenge, and as such a computationally efficient Vecchia process is used as an approximation to the full Gaussian process Katzfuss and Guinness [2021]. The accuracy of the Vecchia approximation for the kernel-convolution framework is demonstrated through comparison results on subsets of the data in Section 2.5. The results of the model fit to the Argo data restricted to January are displayed in Section 2.6, demonstrating its ability to compute confidence and credible intervals for the parameter fields, ocean heat content, and the heat content trend over time. Section 2.7 shows that the proposed model out-performs an out-of-the-box method at predicting ocean heat content through cross-validation, and also shows the relative importance of the various aspects of the model. The functions used in performing this analysis are available as the R package **BayesianOHC**.

2.2 Data

After deployment, an Argo float proceeds through four stages. First, the float adjusts its buoyancy to descend to 1,000m of depth. In the second stage, the float drifts at 1,000m of depth for an average of nine days, after which it descends further to 2,000m. The float then ascends from 2,000m to the surface, measuring temperature, salinity, and pressure at depth intervals averaging between 50 and 100 meters. Once the float surfaces, the measurements as well as the float's location and time at surfacing are transmitted via satellite [Riser et al.,

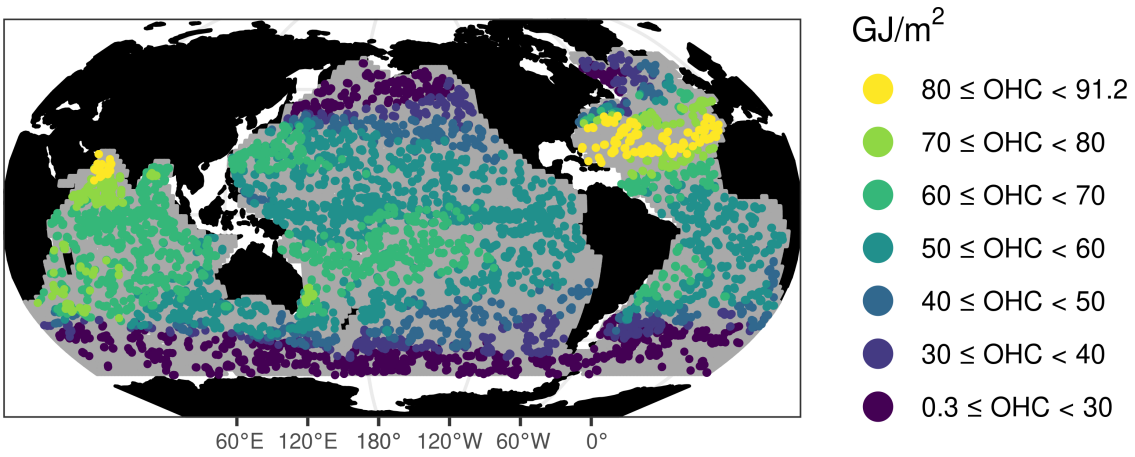


Figure 2.1: Vertical column ocean heat content values measured by Argo floats during January 2016. The mask used as the domain for interpolating ocean heat content is shaded in gray.

2016].

For the purposes of developing a Bayesian model for the spatial process without the additional challenges of modeling spatio-temporal behavior, the analysis here is restricted to profiles observed during January, although the model developed here will be extended to all months in Chapter 3. For the purpose of ignoring temporal correlations, each data point is taken as occurring at the mid-point of the month. Attention is restricted to profiles observed between 2007 and 2016 as this is the range of years included in the Argo data obtained from Kuusela and Stein [2018], although data up to 2021 is incorporated in Chapter 3. For various reasons, some of the observations in the dataset were not recorded to the full depth of 2,000m, and the percentage of floats with a maximum depth greater than 1,900m ranges from 50% in 2007 to 70% in 2016. To avoid having to extrapolate the temperature field over depth, all floats with a maximum recorded depth shallower than 1,900m are excluded from the analysis. In total, the data under consideration consists of 42,776 profiles, ranging from 2,581 in 2007 to 7,543 in 2016.

For notation, let $yr \in \{2007, \ldots, 2016\}$ denote the year, $\mathbf{x} = (x_{\text{lat}}, x_{\text{lon}}) \in \mathbb{S}^2$ denote

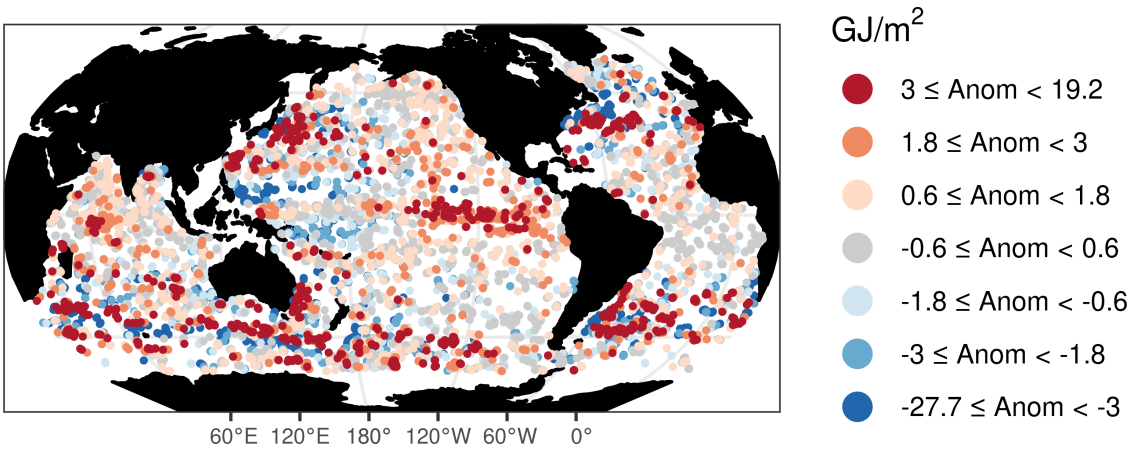


Figure 2.2: Observed heat content anomaly field, where anomalies are computed with respect to the spatially-varying mean-field estimated in section 2.6.

spatial location in latitude-longitude coordinates, and $z \in \mathbb{R}^+$ denote depth from the surface in meters. Define the three-dimensional temperature field at location $\mathbf{x}$, depth z, and year yr at mid-January as $T_{yr}(\mathbf{x}, z)$ measured in degrees Celsius. Since the ultimate interest is in the integrated heat content, the proposed model will be for the two-dimensional spatial field of heat content values integrated over depth rather than the temperature and/or salinity fields at fixed depth levels. The vertically integrated heat content field from 0-2,000m depth is defined as

$$H_{yr}(\mathbf{x}) = \int_{0\text{m}}^{2000\text{m}} \rho c_p T_{yr}(\mathbf{x}, z)\, dz \tag{2.1}$$

where ρ is seawater density in kg/m^3 and c_p is the specific heat of seawater in J/(kg$^\circ$ C) [Dijkstra, 2008]. The units of $H_{yr}(\mathbf{x})$ are J/m^2. For observed temperature profiles the integral in Equation (2.1) is computed numerically by summing the values of the observations linearly interpolated to each meter of the domain [0m,2000m]. The heat content values as observed in January 2016 are shown in Figure 2.1. Heat content anomalies computed with respect to the spatially-varying mean-field estimated in Section 2.6 are shown in Figure 2.2.

The main quantity of interest in this work is the total ocean heat content or OHC, which

is defined for a given year as the two-dimensional integral of the heat content field:

$$\text{OHC}_{yr} = \int_{\mathbf{x} \in S_{\text{mask}}} H_{yr}(\mathbf{x})\, d\mathbf{x}$$

where $S_{\text{mask}} \subset \mathbb{S}^2$ is the ocean's surface masked following Roemmich and Gilson [2009]; this mask excludes regions of the ocean that are hard or impossible to sample with Argo floats, including regions of the ocean where the depth is less than 2,000m, seasonally ice-covered regions, and marginal seas such as the Gulf of Mexico and the Mediterranean. The mask can be seen as the gray background of Figure 2.1. The primary goal of this work is to estimate the trend in OHC_{yr}, and quantify the uncertainty in those estimates, using the observed Argo profiles. Let $H_{\text{obs};yr}(\mathbf{x})$ denote the discrete integral of the temperature profile observed at location $\mathbf{x}$ and year yr. In this proposed modeling framework, the underlying heat content field H_{yr} differs from the observed heat content field $H_{\text{obs};yr}$ in that the latter contains small-scale variability that will be modeled as a nugget effect; see section 2.3.1 for further details. Denote the vector of observed values for year $yr \in \{2007, \ldots, 2016\}$ as $\mathbf{H}_{\text{obs};yr}$ and the concatenated vector of observed values as $\mathbf{H}_{\text{obs}} \equiv [\mathbf{H}_{\text{obs};2007}, \ldots, \mathbf{H}_{\text{obs};2016}]$.

2.3 Model Framework

The proposed approach is to use a hierarchical Bayesian framework for the ocean heat content model, where at the first level of the hierarchy the ocean heat content field is modeled as a Gaussian process, and in the second level, the spatially-varying model parameters are modeled as Gaussian hyper-processes. This approach yields several advantages. First, it provides a framework for propagating and quantifying uncertainty from both the inferred parameter fields and the spatial prediction of ocean heat content, thereby allowing us to determine the relative importance of each type of uncertainty. Second, the parameter fields are allowed to vary smoothly in space according to a Gaussian process with prescribed hyperparameters, without the need to choose a functional form. This flexibility is critical for modeling ocean heat content due to the complex structure of the anomaly field as can

be seen in Figure 2.1 and Figure 2.2; this will be discussed quantitatively in Section 2.7.

This section will describe the non-stationary covariance model used for the heat content field, the hierarchical Bayesian framework including Gaussian process priors on the parameter fields, and the estimation of the posterior distribution of OHC using MCMC sampling with Metropolis-Hastings-within-Gibbs steps.

2.3.1 Covariance Model for the Heat Content Field

The heat content field is modeled as a Gaussian process with spatially-varying covariance parameters in order to capture non-stationarity in the data. In the following, the mean-field will be denoted as μ, the variance field as ϕ, the nugget variance field as σ^2, and the latitudinal and longitudinal correlation length scale fields as θ_{lat} and θ_{lon} respectively. For notation these symbols will refer to the fields as functions of location, so for example the value of the latitudinal correlation length scale at location $\mathbf{x}$ would be $\theta_{\text{lat}}(\mathbf{x})$. The field μ is defined for any $\mathbf{x}$ and year yr as $\mu(\mathbf{x}) = \mu_{2007}(\mathbf{x}) + \beta(\mathbf{x})(yr - 2007)$ where μ_{2007} is a spatially-varying mean-field centered on 2007 and β is a spatially-varying trend field. Denote the collection of parameter fields as $\boldsymbol{\rho} \equiv \{\phi, \theta_{\text{lat}}, \theta_{\text{lon}}, \sigma^2, \mu_{2007}, \beta\}$. The observed and underlying heat content fields are distributed as

$$H_{\text{obs};yr}(\mathbf{x}) \overset{\text{i.i.d.}}{\sim} N(H_{yr}(\mathbf{x}), \sigma^2(\mathbf{x}))$$

and

$$H_{yr}(\mathbf{x}) \sim GP(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{y}; \boldsymbol{\rho})) \tag{2.2}$$

respectively where GP is a Gaussian process such that $k(\mathbf{x}, \mathbf{y}; \boldsymbol{\rho}) \equiv \text{cov}[H_{yr}(\mathbf{x}), H_{yr}(\mathbf{y})]$ for any pair of locations $\mathbf{x}$ and $\mathbf{y}$. Here, for notational simplicity k is written to be dependent on the entire set of parameter fields $\boldsymbol{\rho}$, although the value of k between two locations will depend only on the values of θ_{lat}, θ_{lon}, and ϕ at those locations.

The nugget variance σ^2 is generally interpreted as measurement error, however as measurement errors in Argo temperature readings are known to be small [Riser et al., 2016]

it will be used here to represent fine-scaled variation in the underlying field. This can be induced, for example, by eddies, which have coherent structure but are generally smaller in spatial and temporal scale than the observation resolution of the Argo profiles [Colling, 2001]. This form of variation is excluded from the kriging-based estimates of ocean heat content anomaly in Equation (2.2) but is represented in the uncertainty estimates. In practice the inverse signal-to-noise ratio $\sigma^2(\mathbf{x})/\phi(\mathbf{x})$ is used as this field appears to have better sampling properties, however for simplicity the exposition in this section describes the analogous case where σ^2 is sampled directly.

Kernel convolution methods as introduced by Higdon [1998b] are a straightforward and intuitive way to integrate a spatially-varying parameter field into a valid covariance model. The idea behind these methods is to assign a kernel to each point in the domain that gives the local covariance properties at that location. Then a valid global Gaussian process can be obtained by computing the convolution of the kernels between every two points. The correlation between any two points will reflect the local behavior at both of their locations. Paciorek and Schervish [2006b] demonstrated the use of kernel convolution methods and find closed-form expressions for the covariance functions when the underlying domain is Euclidean. For modeling the ocean heat content field it would make the most sense to measure distances between observations using the great-circle distance, which is the shortest distance between two points over the sphere's surface. Unfortunately, the kernel convolutions do not have closed-form expressions when the great-circle distance is used, and the computational task of performing numerical integration of the convolution terms as done in Li and Zhu [2016b] is not computationally feasible for the number of observations in the Argo data.

In principle, the spherical domain could be represented as a subset of three-dimensional Euclidean space and the chord-length distance, which is the shortest line between two points and is allowed to cross through the interior of the sphere, could be used. However, in $\mathbb{R}^3$ the traditional method of representing anisotropy through scaling and rotating the coordinate space would not yield parameters that could be interpreted in terms of quantities on the

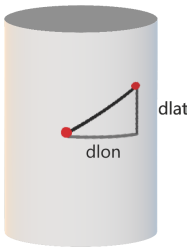


Figure 2.3: Visualization of the cylindrical distance metric. Here, *dlat* is the Euclidean distance between the latitude coordinates and *dlon* is the great-circle distance between the longitude coordinates.

surface. This is a limitation when modeling the ocean heat content field which, as can be seen in Figure 2.2, exhibits significant anisotropy in the latitudinal and longitudinal dimensions. To examine the significance of this limitation, the cross-validation study of Section 2.7 evaluates the ability of isotropic models to predict ocean heat content.

The proposed solution to the problem of modeling anisotropy on the sphere is to represent the Earth's surface in latitude-longitude coordinates as a cylinder, or $\mathbb{CL} = \mathbb{R} \times \mathbb{S}^1$, and use a cylindrical distance function for the covariance model that allows for the estimation of correlation length scales in the latitudinal and longitudinal dimensions. As Argo floats do not travel close to the poles, the approximation of the Earth's surface as a cylinder does not present any modeling problems. In particular, the geometric variation in the radius of the sphere with latitude can be accounted for in the estimation of the spatially-varying longitudinal scale parameter. Let $d_{\text{euc}}(x_{\text{lat}}, y_{\text{lat}})$ denote the Euclidean distance between two points $x_{\text{lat}}, y_{\text{lat}} \in \mathbb{R}$ and let $d_{\text{gc}}(x_{\text{lon}}, y_{\text{lon}})$ denote the great circle distance between two points $x_{\text{lon}}, y_{\text{lon}} \in \mathbb{S}^1$. Then the shortest path between two locations on the cylinder is given by the Pythagorean theorem as visualized in Figure 2.3. Latitudinal and longitudinal correlation length scale parameters can be incorporated in a straightforward way, yielding the following distance function, where $\mathbf{x} = [x_{\text{lat}}, x_{\text{lon}}]$ and $\mathbf{u} = [u_{\text{lat}}, u_{\text{lon}}]$ are arbitrary locations:

$$d_{\text{cyl}}(\mathbf{x}, \mathbf{u}; \theta_{\text{lat}}, \theta_{\text{lon}}) = \sqrt{\frac{d_{\text{euc}}(x_{\text{lat}}, u_{\text{lat}})^2}{\theta_{\text{lat}}(\mathbf{x})} + \frac{d_{\text{gc}}(x_{\text{lon}}, u_{\text{lon}})^2}{\theta_{\text{lon}}(\mathbf{x})}}.$$

Using this distance function in the kernel convolution framework with Gaussian kernels yields the following covariance function for arbitrary locations $\mathbf{x}$ and $\mathbf{y}$:

$$k(\mathbf{x}, \mathbf{y}; \boldsymbol{\rho}) = \sqrt{\phi(\mathbf{x})\phi(\mathbf{y})} \int_{\mathbf{u} \in \mathbb{CL}} \exp\left(-d_{\text{cyl}}(\mathbf{x}, \mathbf{u}; \theta_{\text{lat}}, \theta_{\text{lon}})^2 - d_{\text{cyl}}(\mathbf{y}, \mathbf{u}; \theta_{\text{lat}}, \theta_{\text{lon}})^2\right) d\mathbf{u}. \quad (2.3)$$

By using Gaussian covariance kernels the integral over $\mathbb{CL}$ can be evaluated separately in the latitudinal and longitudinal dimensions. The integral over the Euclidean (latitudinal) dimension can then be computed easily through the closed-form expressions of Paciorek and Schervish [2006b]. The convolutions in the cylindrical dimension can be computed exactly using Gaussian error functions as derived in Section 2.4.1. Section 2.4.2 presents a faster Gaussian approximation to the exact spherical convolutions along with a simulation study demonstrating that the approximation is highly accurate over the correlation length scales present in the ocean heat content data. As such the approximation to the exact convolutions is used in the following analysis, although both the approximation and exact methods are available in the **BayesianOHC** package as "cylindrical_correlation_gaussian" and "cylindrical_correlation_exact" respectively.

What results from Equation (2.3) is a covariance function that can be locally anisotropic and globally non-stationary, is fast to compute, and is mathematically valid on the domain. A major advantage of the cylindrical distance model's anisotropy structure is that the range parameters are directly interpretable as correlation length scales in the latitudinal and longitudinal dimensions. To preserve this interpretability the cylindrical distance model does not allow for rotation of the anisotropy structure.

2.3.2 Hierarchical Bayesian Framework

Having prescribed the structure of the proposed covariance function for ocean heat content, attention can be shifted to the model for the covariance parameters, which are themselves modeled as Gaussian processes. Each parameter field has a stationary Gaussian process prior with mean, variance, and spatial range hyperparameters. Fitting this complex model

is made computationally feasible through the Vecchia process described in Section 2.5.

In the following, let $\Sigma_{\text{obs,obs};\boldsymbol{\rho}}$ denote the covariance matrix between each pair of observations as computed by k. Because the construction here only involves data within discrete and temporally separated months, each year can be treated as independent, and as such $\Sigma_{\text{obs,obs};\boldsymbol{\rho}}$ is a block diagonal matrix. Future work will incorporate temporal dependence structures in the covariance matrix. Let $\rho \in \boldsymbol{\rho}$ denote an arbitrary parameter field. Then the prior for each $\rho \in \boldsymbol{\rho}$ is a stationary Gaussian process with link function g_ρ, specifically

$$g_\rho^{-1}(\rho(\mathbf{x})) \sim GP(\mu_\rho, \phi_\rho \exp(-d_{\text{cyl}}(\mathbf{x}, \mathbf{y})/\theta_\rho))$$

where μ_ρ, ϕ_ρ, and θ_ρ are respectively the mean, variance, and range for the parameter field ρ. The link function g_ρ is exponential for the variance and correlation parameters in order to ensure positivity, and is the identity for the mean and trend parameters. To constrain the number of parameters involved in representing the spatial surfaces, the processes will be represented by independent normal basis values $\mathbf{b}_\rho \sim N(0, I_k)$ on a set of knot locations denoted κ, where n_κ refer to the number of knot points. Specifically, for each ρ define the vector $\mathbf{b}_\rho$ to be such that for any location $\mathbf{x}$ the following holds:

$$\rho(\mathbf{x}) = g_\rho(\mu_\rho + \phi_\rho \Sigma_{\mathbf{x},\kappa;\theta_\rho} \Sigma_{\kappa,\kappa;\theta_\rho}^{-1/2} \mathbf{b}_\rho) \tag{2.4}$$

where $\Sigma_{\mathbf{x},\kappa;\theta_\rho}$ is the covariance vector between $\mathbf{x}$ and the knot locations κ and $\Sigma_{\kappa,\kappa;\theta_\rho}$ is the covariance matrix between points in κ. Intuitively, $\mathbf{b}_\rho$ is a de-meaned and de-correlated representation of $g^{-1}(\rho(\mathbf{x}))$ evaluated at the knot locations, from which the field can be re-constructed through kriging. In the model fit in Section 2.6, κ is chosen to contain knot locations at every 16 degrees in the longitudinal direction and every 8 degrees in the latitudinal direction. Restricting the knot locations with the data mask expressed in Figure 2.1, this results in $n_\kappa = 253$ basis values for each parameter field. This knot selection is somewhat arbitrary, and more principled methods for selecting knot point locations have been explored, such as Katzfuss [2013] who incorporate the knot locations into the Bayesian

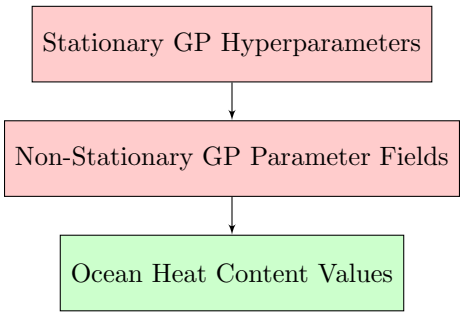


Figure 2.4: Diagram of the hierarchical Bayesian model for the ocean heat content field.

hierarchy through a reversible-jump process, however such approaches are not considered here. The Bayesian hierarchical framework used here is summarized in Figure 2.4.

2.3.3 MCMC Sampling Procedure

To estimate the joint distribution Markov Chain Monte Carlo (MCMC) with Gibbs steps will be used for each of the parameter fields. The marginal distributions for the correlation length and variance parameters must be approximated using Metropolis-Hastings steps, however μ_{2007} and β can be sampled jointly from their Gibbs marginal distribution given by the formula

$$[\mathbf{b}_{\mu_{2007}}, \mathbf{b}_{\beta}] | \mathbf{H}_{\text{obs}}, \mathbf{b}_{\boldsymbol{\rho} \setminus \{\mu_{2007}, \beta\}} \sim N(\mu_{\mu;\text{post}}, \Sigma_{\mu;\text{post}}) \tag{2.5}$$

where

$$\Sigma_{\mu;\text{post}} = (M^T \Sigma^{-1}_{\text{obs,obs};\boldsymbol{\rho}} M + I_{2n_\kappa})^{-1},$$

$$\mu_{\mu;\text{post}} = \Sigma_{\mu;\text{post}} \left(M^T \Sigma^{-1}_{\text{obs,obs};\boldsymbol{\rho}} \mathbf{H}_{\text{obs}} \right),$$

and $M \in \mathbb{R}^{n_{\text{obs}}, 2n_\kappa}$ is the design matrix for $\mathbf{b}_{\mu_{2007}}$ and $\mathbf{b}_{\beta}$. In the conditional distribution of Equation (2.5), "\" is the set minus operator and as such $\mathbf{b}_{\boldsymbol{\rho} \setminus \{\mu_{2007}, \beta\}}$ denotes the set of basis parameters for all of the parameter fields excluding μ_{2007} and β.

Now the MCMC sampling procedure can be outlined. For each iteration $t > 1$ after initialization the following steps are performed:

- For a given $\rho \in \{\theta_{\text{lat}}, \theta_{\text{lon}}, \sigma^2, \phi\}$, propose basis point candidates through independent perturbations, specifically $\mathbf{b}_\rho^{(t)} \sim N\left(\mathbf{b}_\rho^{(t-1)}, \left(\sigma_{\rho;\text{prop}}^{(t)}\right)^2 I_{n_\kappa}\right)$ where $\sigma_{\rho;\text{prop}}^{(t)}$ is the proposal standard deviation at the current iteration. The independent perturbations on $\mathbf{b}_\rho$ can be viewed as perturbing $\log \rho^{(t)}$ by a spatially correlated field that shares the spatial range of $\log \rho^{(t)}$.

- Accept or reject the entire basis vector, which describes the full spatial field, using the Metropolis-Hastings acceptance ratio on

$$P(\mathbf{b}_\rho^{(t)} | \mathbf{H}_{\text{obs}}, \mathbf{b}_{\boldsymbol{\rho}\setminus\{\rho\}}) \propto P(\mathbf{H}_{\text{obs}} | \rho^{(t)}(\mathbf{x}_{\text{obs}}), \mathbf{b}_{\boldsymbol{\rho}}) P(\mathbf{b}_\rho^{(t)})$$

  where $\mathbf{b}_{\boldsymbol{\rho}\setminus\{\rho\}}$ denotes the set of basis points corresponding to all of the fields except for ρ at their most recent iteration and $\mathbf{x}_{\text{obs}}$ denotes the full set of observation locations. The values of ρ at the observation locations can be evaluated through Equation 2.4. Here, the data likelihood term on the right is multivariate Gaussian with covariance matrix $\Sigma_{\text{obs,obs};\boldsymbol{\rho}}$ and the prior term $P(\mathbf{b}_\rho^{(t)})$ is the product of independent standard normal densities.

- Repeat the previous two steps for sampling $\mathbf{b}_\rho^{(t)}$ for each remaining $\rho \in \{\theta_{\text{lat}}, \theta_{\text{lon}}, \sigma^2, \phi\}$.

- Sample $\mathbf{b}_{\mu_{2007}}^{(t)}$ and $\mathbf{b}_\beta^{(t)}$ jointly from their conditional distribution given in Equation (2.5).

- Update the variances $\sigma_{\rho;\text{prop}}^{(t)}$ using the adaptive MCMC algorithm for Metropolis-within-Gibbs sampling as described in Roberts and Rosenthal [2009], which updates the proposal variances to achieve a desired acceptance rate chosen to be .44.

Through this approach, MCMC samples of coherent spatial fields for each model parameter can be obtained. In particular, the mean-field parameters are modeled jointly with the covariance parameters, which will generally lead to improved estimates of both. Further, the resulting credible intervals will incorporate the uncertainty induced by estimating the mean

and the spatially-varying trend. Quantifying the trend is particularly valuable as for many purposes the primary concern is with the trend in ocean heat content anomaly over time rather than the values of the anomalies themselves.

2.3.4 Posterior Distributions

For a fixed set of parameter values sampled from the posterior, the distribution of the globally-integrated ocean heat content, OHC_{yr}, is normal and can be denoted $\text{OHC}_{yr}|\boldsymbol{\rho}, \mathbf{x}_{yr;\text{obs}} \sim N(\mu_{\text{OHC}_{yr}}, \sigma^2_{\text{OHC}_{yr}})$ with values

$$\mu_{\text{OHC}_{yr}} = \Sigma_{\text{globe,obs};\boldsymbol{\rho}} \Sigma_{\text{obs,obs};\boldsymbol{\rho}} \mathbf{H}_{\text{obs};yr} \text{ and}$$

$$\sigma^2_{\text{OHC}_{yr}} = \Sigma_{\text{globe,globe};\boldsymbol{\rho}} - \Sigma_{\text{globe,obs};\boldsymbol{\rho}} \Sigma^{-1}_{\text{obs,obs};\boldsymbol{\rho}} \Sigma_{\text{obs,globe};\boldsymbol{\rho}} \tag{2.6}$$

where $\Sigma_{\text{globe,obs};\boldsymbol{\rho}}$ denotes the covariance vector between the observations locations and the global integral and $\Sigma_{\text{globe,globe};\boldsymbol{\rho}}$ denotes the variance of the global integral. When $\mu_{\text{OHC}_{yr}}$ and $\sigma_{\text{OHC}_{yr}}$ are computed from the sampled parameters $\boldsymbol{\rho}^{(t)}$ they will be referred to as $\mu^{(t)}_{\text{OHC}_{yr}}$ and $\sigma^{(t)}_{\text{OHC}_{yr}}$. The entries in these covariance matrices involve integrals over the global domain which must be computed numerically. Let $\mathbf{x}_{\text{grid}}$ denote the grid-points to be used for numeric integration; then the approximation of ocean heat content can be written as $\text{OHC}_{yr} = a^T_{\text{globe}} \mathbf{H}_{\text{globe};yr}$ where $\mathbf{H}_{\text{globe};yr}$ are the unknown ocean heat content values at the grid-points and a_{globe} is the vector of grid-cell areas corresponding to $\mathbf{x}_{\text{globe}}$. Then the above formulas apply replacing $\Sigma_{\text{globe,globe}}$ with $\Sigma_{\text{grid,grid}}$. In the results reported in Section 2.6, OHC posterior distributions are calculated using a grid partitioning the domain with a resolution of $1^\circ \times 1^\circ$.

2.4 Cylindrical Kernel Convolutions

This section will outline how the cylindrical kernel convolutions can be computed exactly using Gaussian error functions in the spherical dimension. Then a simulation study will

demonstrate that a Gaussian function approximation to the exact convolutions is accurate over the range of longitudinal correlations present in the Argo data.

2.4.1 Formulas for Exact Convolutions

The kernel convolution covariance described in Equation (2.3) in Section 2.3.1 can be written as

$$k(\mathbf{x},\mathbf{y};\boldsymbol{\rho}) = \phi(\mathbf{x})\phi(\mathbf{y})\int_{\mathbf{u}\in\mathbb{CL}} \exp\left(-d_{\text{cyl}}(\mathbf{x},\mathbf{u};\theta_{\text{lat}},\theta_{\text{lon}})^2 - d_{\text{cyl}}(\mathbf{y},\mathbf{u};\theta_{\text{lat}},\theta_{\text{lon}})^2\right)\, d\mathbf{u}$$

$$= \int_{\mathbb{R}} \exp\left(\frac{d_{\text{euc}}(x_{\text{lat}},u_{\text{lat}})^2}{\theta_{\text{lat}}(x_{\text{lat}})} - \frac{d_{\text{euc}}(y_{\text{lat}},u_{\text{lat}})^2}{\theta_{\text{lat}}(y_{\text{lat}})}\right) du \int_{\mathbb{S}^1} \exp\left(\frac{d_{\text{gc}}(x_{\text{lon}},u_{\text{lon}})^2}{\theta_{\text{lon}}(x_{\text{lon}})} - \frac{d_{\text{gc}}(y_{\text{lon}},u_{\text{lon}})^2}{\theta_{\text{lon}}(y_{\text{lon}})}\right) du$$

Let $G_a^b(x,y;\theta_x,\theta_y) = \int_a^b \exp\left(-\frac{(u-x)^2}{\theta_x} - \frac{(u-y)^2}{\theta_y}\right) du$ denote the Gaussian convolution with generic parameters and limits of integration. Through completing the square it can be seen that

$$G_a^b(x,y;\theta_x,\theta_y) = \exp\left(\frac{(y-x)^2}{\theta_x+\theta_y}\right)\int_a^b \frac{u - \frac{x\theta_x+y\theta_y}{\theta_x+\theta_y}}{\theta_x\theta_y/(\theta_x+\theta_y)}\, du$$

$$= \sqrt{\pi}\sqrt{\theta_{xy}}\exp\left(\frac{(y-x)^2}{\theta_x+\theta_y}\right)\left[\Phi\left(b;\frac{x\theta_x+y\theta_y}{\theta_x+\theta_y},\sqrt{\theta_{xy}}\right) - \Phi\left(a;\frac{x\theta_x+y\theta_y}{\theta_x+\theta_y},\sqrt{\theta_{xy}}\right)\right] \quad (2.7)$$

where $\theta_{xy} = \frac{\theta_x\theta_y}{\theta_x+\theta_y}$ and Φ denotes the normal cumulative density function (CDF) with mean and variance given as the first and second arguments respectively. By taking the limits of Equation (2.7) it can be seen that the latitudinal dimension term of the kernel convolution has a simple closed-form expression:

$$\int_{\mathbb{R}} \exp\left(\frac{d_{\text{euc}}(x_{\text{lat}},u_{\text{lat}})^2}{\theta_{\text{lat}}(x_{\text{lat}})} - \frac{d_{\text{euc}}(y_{\text{lat}},u_{\text{lat}})^2}{\theta_{\text{lat}}(y_{\text{lat}})}\right) du = G_{-\infty}^{\infty}(x_{\text{lat}},y_{\text{lat}};\theta_{\text{lat}}(x_{\text{lat}}),\theta_{\text{lat}}(y_{\text{lat}}))$$

$$= \sqrt{\pi}\sqrt{\frac{\theta_{\text{lat}}(x_{\text{lat}})\theta_{\text{lat}}(y_{\text{lat}})}{\theta_{\text{lat}}(x_{\text{lat}})+\theta_{\text{lat}}(y_{\text{lat}})}}\exp\left(-\frac{1}{2}\frac{(x_{\text{lat}}-y_{\text{lat}})^2}{\theta_{\text{lat}}(x_{\text{lat}})+\theta_{\text{lat}}(y_{\text{lat}})}\right) \quad (2.8)$$

For the cylindrical integral the locations x and y are first rotated to $x' = 0$ and $y' = d_{\text{gc}}(x,y)$ respectively, a transformation which preserves the great-circle distance between the two points. Since $u \in [-\pi,\pi]$ the first distance term becomes $d_{\text{gc}}(x',u) = (x'-u)^2$. For

the second distance term, the integral can be split between $u \in (y-\pi, \pi)$ where $d_{\text{gc}}(y', u) = (y'-u)^2$ and $u \in (-\pi, y-\pi)$ where $d_{\text{gc}}(x', u) = (y'-2\pi-u)^2$. This yields

$$\int_{-\pi}^{\pi} \exp\left(-\frac{d_{\text{gc}}(x', u)}{\theta_x} - \frac{d_{\text{gc}}(y', u)}{\theta_y}\right) du$$

$$= \int_{-\pi}^{y'-\pi} \exp\left(-\frac{(x'-u)^2}{\theta_x} - \frac{(y'-2\pi-u)^2}{\theta_y}\right) du + \int_{y'-\pi}^{\pi} \exp\left(-\frac{(x'-u)^2}{\theta_x} - \frac{(y'-u)^2}{\theta_y}\right) du$$

$$= G_{-\pi}^{u'-\pi}(x', y'-2\pi; \theta_x, \theta_y) + G_{y'-\pi}^{\pi}(x', y'; \theta_x, \theta_y). \tag{2.9}$$

Then by substituting in x_{lon} and $\theta_{\text{lon}}(x_{\text{lon}})$ into the above equation the spherical convolution can be computed using two evaluations of Equation (2.7). This is a closed form expression involving the use of the Gaussian CDF, which has pre-computed evaluation functions in most programming languages. The G function described above is available in the **BayesianOHC** package as the function "gaussian_convolution" and the exact convolutions in the longitudinal and latitudinal dimensions are available as "cylindrical_correlation_exact" and "euclidean_correlation_convolution" respectively. It should be noted that the implementation follows Paciorek and Schervish [2006b] and scales the convolutions by $\frac{\sqrt{2}}{\theta_{\text{lon}}(x_{\text{lon}})^{1/4}\theta_{\text{lon}}(x_{\text{lon}})^{1/4}\sqrt{\pi}}$ and $\frac{\sqrt{2}}{\theta_{\text{lat}}(x_{\text{lat}})^{1/4}\theta_{\text{lat}}(x_{\text{lat}})^{1/4}\sqrt{\pi}}$ respectively to ensure that the correlation between a location and itself evaluates to one.

2.4.2 Simulation Study Justifying the Use of Gaussian Approximation

While Gaussian CDFs are faster to compute than estimating the integrals in the convolution numerically, evaluating the pre-computed CDF still takes longer to compute than the latitudinal convolutions using Equation (2.8). One would expect that when the correlation length scales are small relative to the circumference of the Earth, the circular convolutions will behave similarly to the Euclidean convolutions and the following alteration of Equation (2.8) could be used to approximate the convolution:

$$= \sqrt{\pi}\sqrt{\frac{\theta_{\text{lon}}(x_{\text{lon}})\theta_{\text{lon}}(y_{\text{lon}})}{\theta_{\text{lon}}(x_{\text{lon}}) + \theta_{\text{lon}}(y_{\text{lon}})}} \exp\left(-\frac{1}{2}\frac{d_{\text{gc}}(x_{\text{lon}}, y_{\text{lon}})^2}{\theta_{\text{lon}}(x_{\text{lon}}) + \theta_{\text{lat}}(y_{\text{lon}})}\right). \tag{2.10}$$

In order to determine the size of the correlation lengths that would be required for Equation (2.10) to be a sufficient approximation for the exact convolutions, a simulation was performed to compare the two correlation functions on data randomly generated according to known parameters. First, one hundred locations x were simulated from the uniform distribution of the unit circle. Then a dense grid of θ was generated corresponding to an effective range length between zero and 120° degrees, where "effective range" here is the distance at which the correlation becomes less than 0.05. For each value of θ two types of correlation structures were simulated: a stationary structure where $\theta(x) = \theta$ for all locations x and a non-stationary structure $\theta(x) = \cos(x)\theta/2+\theta$. In the non-stationary structure, local range parameters vary between $\theta/2$ and $3\theta/2$ with θ being the mid-point of the parameter field. For each value of θ one hundred sets of observations were randomly generated using covariance matrices generated from the exact convolution method for both the stationary and non-stationary correlation structures. Then for each simulated set of data, maximum likelihood estimates (MLEs) were computed according to both the exact method and the approximation method, and the average was computed over the 100 simulations. Finally fractional errors of the approximation method MLEs over the exact method MLEs where calculated, specifically $\frac{\hat{\theta}_{\text{approx}}-\hat{\theta}_{\text{exact}}}{\hat{\theta}_{\text{exact}}}$ where $\hat{\theta}$ refers to the MLEs averaged over the one hundred simulations. Figure 2.5 shows the fractional errors for the stationary and non-stationary correlation structures over each value of θ displayed in terms of the effective range in degrees.

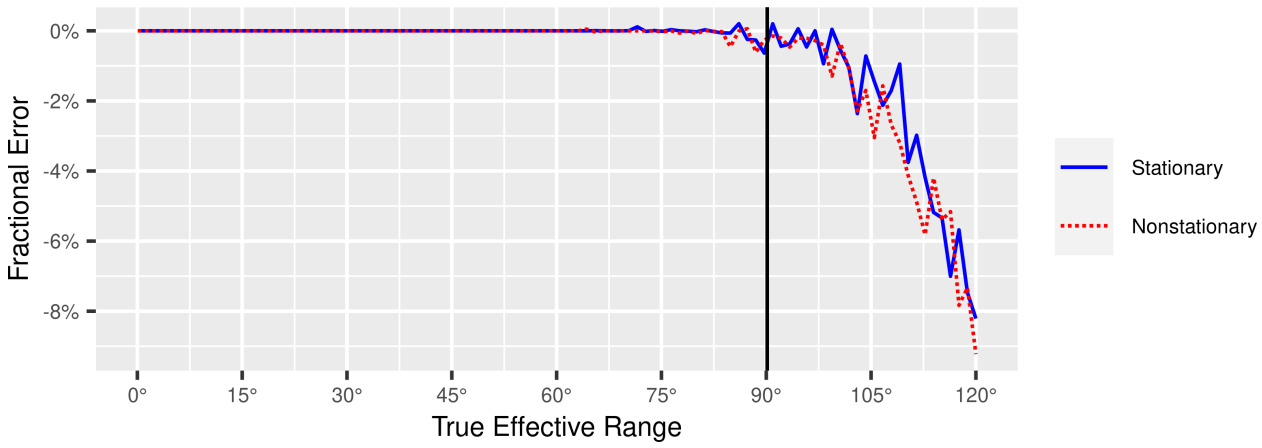


Figure 2.5: Fractional errors of the exact method over the Gaussian approximation. The horizontal line is at the 99$^{\text{th}}$ percentile of the initial condition distribution of θ_{lon}.

It can be seen that the approximation is nearly exact for effective range parameters less than 80°, beyond which the approximation method begins to yield increasingly larger negative errors. This trend is the same between the stationary and non-stationary simulations. The fact that the approximation does increasingly worse at high values of the range parameter is to be expected, as at this point the value of the exact convolution will begin to diverge from the Gaussian approximation.

The horizontal line in Figure 2.5 shows the 99$^{\text{th}}$ percentile of θ_{lon} in the fitted fully non-stationary cylindrical distance model, which corresponds to an effective range of 90.13. At this value, the stationary and non-stationary fractional errors are $-.6\%$ and $-.2\%$ respectively. The small error of the approximation at the largest correlation length scales present in the ocean heat content field demonstrates that the Gaussian approximation to the exact spherical convolutions is sufficient for modeling ocean heat content. It should be noted that at the initial configuration used for the MCMC configuration, the 99$^{\text{th}}$ quantile of the effective ranges was 65.40° at which point the fractional errors from the simulation are less than a hundredth of a percent. Due to its sufficiency and computational advantages, the approximation will be used in the following results, and is available in the **BayesianOHC**

package as the function "cylindrical_correlation_gaussian".

2.5 Vecchia Processes for Feasible Bayesian Estimation

The most computationally intensive step of the Metropolis-Hastings procedure is computing the log-likelihood term. The log-likelihood term can be written as

$$\log P(\mathbf{H}_{\text{obs}}|\boldsymbol{\rho}) \propto -\left(\mathbf{H}_{\text{obs}} - \mu(\mathbf{x}_{\text{obs}})\right)^T \Sigma_{\text{obs,obs}}^{-1} \left(\mathbf{H}_{\text{obs}} - \mu(\mathbf{x}_{\text{obs}})\right) - \log\det(\Sigma_{\text{obs,obs}}^{-1})$$

where the dependency of $\Sigma_{\text{obs,obs}}$ on $\boldsymbol{\rho}$ is omitted for notational simplicity. The usual way to compute the inverse and log-determinant terms in the above equation is to first compute the Cholesky decomposition of the covariance matrix, resulting in the upper-triangular Cholesky factor C which satisfies the property $C^T C = \Sigma_{\text{obs,obs}}$. Then the likelihood evaluation becomes

$$\log P(\mathbf{H}_{\text{obs}}|\boldsymbol{\rho}) \propto -\left(C^{-1}(\mathbf{H}_{\text{obs}} - \mu(\mathbf{x}_{\text{obs}})\right)^T \left(C^{-1}(\mathbf{H}_{\text{obs}} - \mu(\mathbf{x}_{\text{obs}}))\right) - 2\sum_{i=1}^{n} \log C_{ii}. \quad (2.11)$$

Inverting an upper triangular matrix is computationally simple, so the primary burden is computing the Cholesky factorization. This step has a computational complexity of $O(n^3)$, meaning that the computation time increases cubically with the number of observations. For example, in a simple experiment on a standard laptop computer, computing the likelihood for the 2,581 Argo observations in 2007 takes around 7 seconds, while computing the likelihood for the 7,542 observations in 2016 takes around 183 seconds. While a single likelihood evaluation is feasible, since each full Metropolis-Hasting step requires four covariance matrix inversions it would take over five months to compute 20,000 samples using the Cholesky method even when computing each year's likelihood in parallel.

2.5.1 Vecchia Approximation

To render Bayesian estimation feasible the Vecchia approximation of Vecchia [1988] is used. This technique approximates a Gaussian process by writing the likelihood as a series of con-

ditional likelihoods and then restricting the set of observations used by the conditional likelihoods. Specifically, let $H_{\text{obs},1}, \ldots, H_{\text{obs},n_{\text{obs}}}$ denote the set of observations and let $g(1), \ldots, g(n_{\text{obs}})$ denote sets of integers such that $g(1) = \varnothing$ and $g(j) \subseteq \{1, \ldots, j-1\}$. When $g(j)$ is more than one integer, $H_{\text{obs},g(j)}$ will denote the set of observations corresponding to the indices in $g(j)$. With this the Vecchia likelihood is defined as

$$\ell_{\text{Vecchia}}(H_{\text{obs}}) = \sum_{j=1}^{n_{\text{obs}}} \ell(H_{\text{obs},j} | H_{\text{obs},g(j)}) \tag{2.12}$$

where $\ell(H_{\text{obs},j} | H_{\text{obs},g(j)})$ is the conditional log-likelihood of $H_{\text{obs},j}$ given $H_{\text{obs},g(j)}$. When $g(j) = \{1, \ldots, j-1\}$, each observation is conditioned on all of the observations preceding it in the ordering, and the Vecchia log-likelihood is equivalent to the Gaussian process log-likelihood. When $g(j)$ is a strict subset of the preceding indices, then the Vecchia likelihood is an approximation. The computation of each $\ell(H_{\text{obs},j} | H_{\text{obs},g(j)})$ term requires computing the Cholesky factorization for observations within $g(j)$, so computational gains are achieved when the $g(j)$ are chosen to be small. Generally, the conditioning sets are chosen to be such that $|g(j)| < m$ for some $m << n_{\text{obs}}$ so that the computation time of the Vecchia likelihood is $O(nm^3)$, as opposed to $O(n^3)$ for the full likelihood. The parameter m is tunable, with larger values yielding a more accurate approximation and smaller values enabling faster computation times. An additional advantage of the Vecchia likelihood is that the terms in Equation (2.12) can be computed in parallel whereas parallelization cannot be used directly when computing the Cholesky factorization for the full covariance matrix.

There are three steps for constructing the Vecchia process for a particular set of observation locations. The first step is to order the locations, which is important because the conditioning sets for a particular observation can only include observations that precede it in the ordering. This step uses the max-min distance ordering advocated for by Guinness [2018]. The second step is to construct the conditioning sets such that each set has no more than m observations. The conditioning sets are taken to be the observation's m-nearest neighbors under the logic that nearby observations are the most influential and as such will

yield the closest approximation to the Cholesky likelihood. The third step is to group the observations whose conditioning sets have substantial overlap. Computing the conditional likelihoods on the grouped observations has been shown by Guinness [2018] to increase the accuracy of the approximation while simultaneously decreasing the computation time. These three steps must be done separately for each year of the Argo data, as the observation locations differ between the years. However, since the conditioning sets and groups do not depend on the parameters, the Vecchia construction needs to only be done once before the beginning of the sampling procedure.

For a particular ordering, conditioning configuration, and grouping, it can be shown that the Vecchia likelihood induces a valid stochastic process on the domain of observation locations [Datta et al., 2016]. This "Vecchia process" can be seen as an approximation of the full Gaussian process with more computationally efficient likelihood evaluations. Under this view the Bayesian inference procedure described in Section 2.3 can be used with the Vecchia process by simply substituting the full Cholesky likelihood with the Vecchia likelihood. Furthermore, the inverse Cholesky decomposition of the covariance matrix implied by the Vecchia process can be constructed directly as a sparse matrix. To see this, following the notation of Katzfuss and Guinness [2021] the summands in Equation (2.12) can be rewritten as $\ell(H_{\text{obs},j}|H_{\text{obs},g(j)}) = \mathcal{N}(H_{\text{obs},j}|B_j H_{\text{obs},g(j)}, D_j)$, where $\mathcal{N}$ indicates the normal density, $B_j = k(\mathbf{x}_{\text{obs},j}, \mathbf{x}_{\text{obs},g(j)})k(\mathbf{x}_{\text{obs},g(j)}, \mathbf{x}_{\text{obs},g(j)})^{-1}$, and $D_j = k(\mathbf{x}_{\text{obs},j}, \mathbf{x}_{\text{obs},j}) - B_j k(\mathbf{x}_{\text{obs},g(j)}, \mathbf{x}_{\text{obs},j})$. Then the elements of the inverse Cholesky factor $\mathbf{U} \in \mathbb{R}^{n_{\text{obs}} \times n_{\text{obs}}}$ are $\mathbf{U}_{jj} = D_j^{-1/2}$ for all j, $\mathbf{U}_{ij} = -(B_j)_{i,g(j)}^T D_j^{-1/2}$ for $i \in g(j)$, and $\mathbf{U}_{ij} = 0$ for $i \notin g(j)$. The sparse matrix $\mathbf{U}$ can be used in place of C^{-1} in Equation (2.11) to compute the Vecchia likelihood. This is no more computationally intensive than computing the sum in Equation (2.12) due to $\mathbf{U}$'s known sparsity structure determined by the conditioning sets.

It can be seen from the construction of $\mathbf{U}$ that for any two observations $j < i$, when $j \notin g(i)$ the partial correlation between $H_{\text{obs},j}$ and $H_{\text{obs},i}$ will be zero, or in other words $H_{\text{obs},j}$ and $H_{\text{obs},i}$ will be conditionally independent. As such the Vecchia process can be

seen as an alteration of the full Gaussian process where some conditional dependencies are removed. It should be noted that this does not necessarily mean that $H_{\text{obs},j}$ and $H_{\text{obs},i}$ are independent, as there could be a path of conditional dependencies between observations through the conditioning sets. As such the nearest-neighbors method of constructing the conditioning sets does not simply eliminate dependencies between distant observations, as is done with other likelihood approximation methods such as covariance tapering. This is important to note in the non-stationary cylindrical model since the Vecchia process is used to estimate spatially-varying correlation length scales, which could potentially be biased if only dependencies between nearest neighbors were taken into account in the likelihood.

2.5.2 Predictions

As discussed in section 2.3.4, the estimate of globally integrated heat content is calculated as the area-weighted sum of unobserved ocean heat content values at grid-points along a grid densely partitioning the domain. This is written as $\text{OHC}_{yr} = a_{\text{globe}}^T \mathbf{H}_{\text{globe};yr}$, where $\mathbf{H}_{\text{globe};yr}$ are the unknown heat content values at the grid-points and a_{globe} are the areal weights of the grid-cells. For sufficiently dense grids, computing the joint conditional distribution over the grid-points as in Equation (2.6) is computationally infeasible. As such, in order to compute the posterior distribution of heat content, it is essential to be able to efficiently compute the joint posterior distribution for a large number of observations. Fortunately, this can be done by augmenting the Vecchia process described above to include the grid-points at which the heat content distribution will be predicted. To do this the "observation-first" ordering in Katzfuss et al. [2020b] is used, which involves ordering the prediction locations $H_{\text{pred},n_{\text{obs}}+1}, \ldots, H_{\text{pred},n_{\text{obs}}+n_{\text{grid}}}$, appending them to the ordering of the observation locations, and then computing the nearest-neighbor conditioning sets for the entire set of indexed locations $1, \ldots, n_{\text{obs}} + n_{\text{grid}}$. Since conditioning sets can only contain preceding observations, the conditioning sets of the observations remain the same while the conditioning sets of prediction locations can contain any observation location as well as any prediction location

preceding it in the ordering.

Recall that $\mathbf{U}$ is the inverse Cholesky factor for the Vecchia process, and let $\mathbf{W} = \mathbf{U}_{\text{pred,all}}^T \mathbf{U}_{\text{all,pred}}$. Then it can be shown that the distribution of the prediction values conditioned on the observations is $\mathbf{H}_{\text{pred}}|\mathbf{H}_{\text{obs}} = N(-\mathbf{W}^{-1}\mathbf{U}_{\text{pred,all}}\mathbf{U}_{\text{all,obs}}\mathbf{H}_{\text{obs}}, \mathbf{W}^{-1})$. It follows that the posterior distribution of ocean heat content given the observations is

$$\text{OHC}_{yr}|\mathbf{H}_{\text{obs};yr} \sim N(-a_{\text{globe}}^T \mathbf{W}^{-1}\mathbf{U}_{\text{pred,all}}\mathbf{U}_{\text{all,obs}}\mathbf{H}_{\text{obs};yr}, a_{\text{globe}}^T \mathbf{W}^{-1} a_{\text{globe}}). \tag{2.13}$$

As $\mathbf{U}$, and subsequently $\mathbf{W}$, are sparse, the terms $\mu_{\text{OHC}_{yr}} \equiv -a_{\text{globe}}\mathbf{W}^{-1}\mathbf{U}_{\text{pred,all}}\mathbf{U}_{\text{all,obs}}\mathbf{H}_{\text{obs};yr}$ and $\sigma_{\text{OHC}_{yr}} \equiv a_{\text{globe}}^T \mathbf{W}^{-1} a_{\text{globe}}$ are both fast to compute. If the full Gaussian process was used to compute this posterior distribution using Equation (2.6), every element of the dense covariance matrix of grid-points would need to be evaluated, which is intractable for sufficiently dense grids.

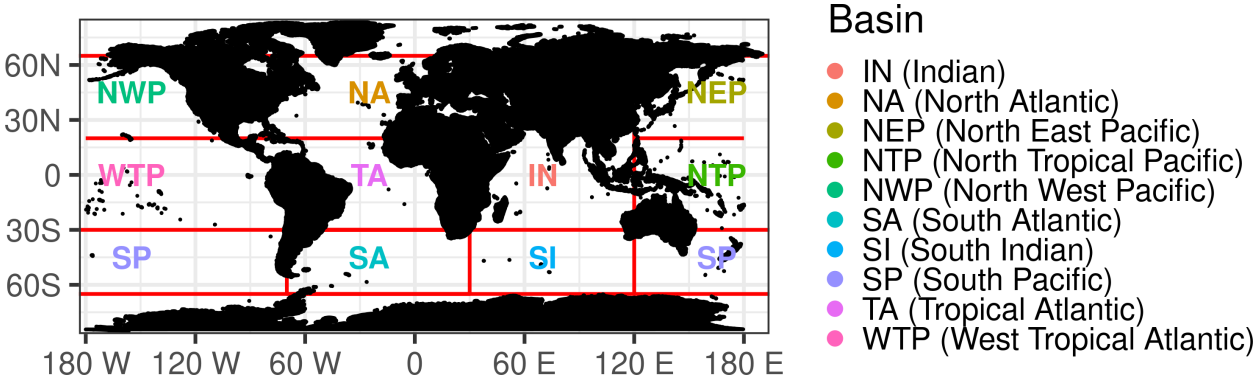


Figure 2.6: Basin definitions used for evaluation of the accuracy of the Vecchia approximation.

2.5.3 Evaluating the Accuracy of the Vecchia Process

The accuracy of the Vecchia likelihood as an approximation to the log-likelihood depends on the ordering, conditioning sets, and grouping used in the construction. In particular, the maximum size of the conditioning sets, determined by the parameter m, needs to be chosen to balance the trade-off between accuracy and computation time. In order to assess the

accuracy of the Vecchia process for the fully non-stationary cylindrical model for different values of m, the global ocean is divided into nine basins. These basins are small enough such that the likelihoods and heat content predictions can be computed easily using the full Cholesky decomposition so that the results from the Cholesky and Vecchia methods can be compared. The basins are defined using the latitudinal and longitudinal limits displayed in Figure 2.6. It should be noted that these basins are for evaluating the approximation accuracy only and that the results in Section 2.6 are computed on the global ocean.

To evaluate the accuracy of the Vecchia approximation on the ocean heat content field, the model described in Section 2.3 is fit independently for each of the nine basins using both the Cholesky method and the Vecchia approximation method for each of $m \in \{10, 25, 50, 100\}$. Using the resulting posterior distributions, the mean, trend, and anomaly fields were interpolated for each basin independently, and globally integrated values were computed using the results within each basin. For the Vecchia process with conditioning parameter m the globally integrated values of the mean, trend, and anomaly fields will be denoted as $\mu_{2007}^{\text{vecc}(m)}$, $\beta^{\text{vecc}(m)}$, and $anom_{yr}^{\text{vecc}(m)}$ respectively. Analogously the globally integrated values for the the Cholesky method will be referred to as μ_{2007}^{chol}, β^{chol}, and $anom_{yr}^{\text{chol}}$. By separating the mean, trend, and anomaly fields the degree to which the Vecchia process accurately approximates these three components can be evaluated. The error in the mean-field would be expected to be the smallest, as the mean-field is highly informed by the values in the data and is not particularly sensitive to the estimated correlation structure. On the other hand, the anomaly field would be expected to have the largest errors, as the predicted anomaly values at unobserved locations are highly sensitive to the correlation structure.

Treating the Cholesky values as "truth", fractional errors $\frac{|\mu_{2007}^{\text{vecc}(m)} - \mu_{2007}^{\text{chol}}|}{\mu_{2007}^{\text{chol}}}$, $\frac{|\beta^{\text{vecc}(m)} - \beta^{\text{chol}}|}{\beta^{\text{chol}}}$, and $\frac{|anom_{yr}^{\text{vecc}(m)} - anom_{yr}^{\text{chol}}|}{|anom_{yr}^{\text{chol}}|}$ are displayed in Table 2.1 for each value of m. For the fractional error of the anomaly fields the median over years is displayed. For $m = 10$ and $m = 25$, the mean-field has smaller fractional errors than the trend or anomaly fields, as expected. The accuracy of the mean-field does not improve noticeably when increasing m beyond 25,

m	mean-field	trend field	anomaly field
10	8.078×10^{-4}	.05029	.7421
25	7.764×10^{-5}	1.400×10^{-3}	.4081
50	2.987×10^{-5}	5.498×10^{-5}	1.665×10^{-3}
100	1.87×10^{-5}	1.963×10^{-5}	1.656×10^{-6}

Table 2.1: Fractional errors for the mean, trend, and anomaly fields of the Vecchia approximation as measured against the analogous values computed using the Cholesky method. Global fields were constructed from results obtained from fitting the model to each of the nine basins in Figure 2.6 independently.

however the error at that level is acceptably small. The approximation error of the trend and anomaly fields are relatively more substantial at $m = 10$, with fractional errors of 5% and 74% respectively. They improve only marginally when increasing m to 25 but see a more substantial improvement when increasing to $m = 50$, with errors around five one-thousandth of a percent and a tenth of a percent respectively. When further increasing to $m = 100$, the approximation errors for the mean and trend fields do not meaningfully change, although the error for the anomaly field does noticeably improve. As the fractional error in the trend field does not decrease after $m = 50$, and at $m = 50$ the level of error in the anomaly fields is acceptable, $m = 50$ will be used for the Vecchia process in the global model fit.

2.6 Results

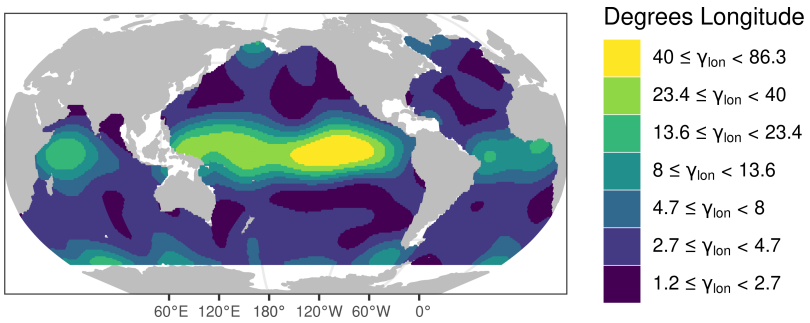
2.6.1 Initial Configuration

As the number of parameters involved in the proposed model is large, it is particularly important to initialize the MCMC sampler at a "good" initial configuration for the sampler to converge in a reasonable number of iterations. To obtain a suitable initial configuration,

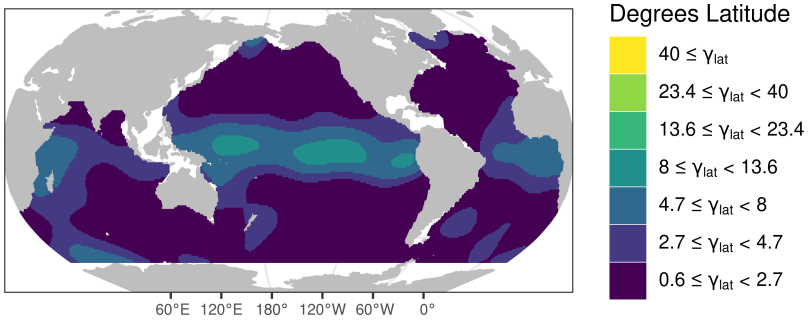
a moving-window estimation method was used analogous to that developed by Kuusela and Stein [2018] to obtain values for the spatially-varying parameter fields. In this approach, the domain is partitioned according to a grid, and at each grid-point, the observations within a window centered around that point are considered to follow a stationary Gaussian process with process variance, nugget variance, latitudinal range, longitudinal range, mean, and trend parameters. Since each of these windows contains a relatively small number of observations, the locally stationary parameters can be estimated quickly using maximum likelihood estimation with Cholesky factorizations.

Parameters were estimated using windows of size $20^\circ \times 20^\circ$ centered at each grid-point of a $6^\circ \times 6^\circ$ grid. Note that the $6^\circ \times 6^\circ$ grid used for the moving window parameter estimates is finer than the $8^\circ \times 16^\circ$ resolution of knot points used in the global model; this is done so that the hyperparameters of the parameter fields can be estimated. To estimate the hyperparameters, stationary and isotropic Gaussian processes were fit to the moving window output for each of the parameter fields, obtaining an estimate of the hyper range, mean, and variance for each process. As described in Section 2.3, the correlation length and variance parameters are assumed to follow a log-normal Gaussian distribution. The hyper-mean of the trend field β is constrained to be zero to ensure that the estimate is fully informed by the data. Additionally, the mean and trend fields were assumed to have the same correlation length scale parameter. The results are shown in Table 2.2, where the hyper-prior distributions are summarized using quantiles since the log-normally distributed priors are non-symmetric. The range hyperparameter for each process is shown in the second column as the "effective range", which is defined as the cylindrical distance in degrees at which the correlation is equal to 0.05, and is denoted as γ_{lat} and γ_{lon} for effective latitudinal and longitudinal range respectively.

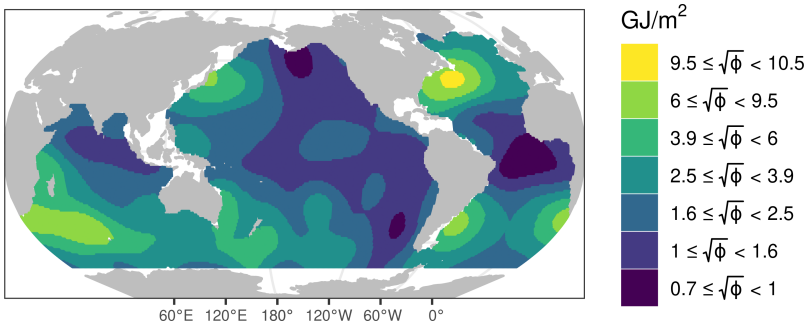
The initial configuration for the MCMC sampler was obtained by kriging the estimated parameters from the moving window approach onto the knot locations using the hyperparameters described in Table 2.2. Figure 2.7 displays the initial configuration interpolated



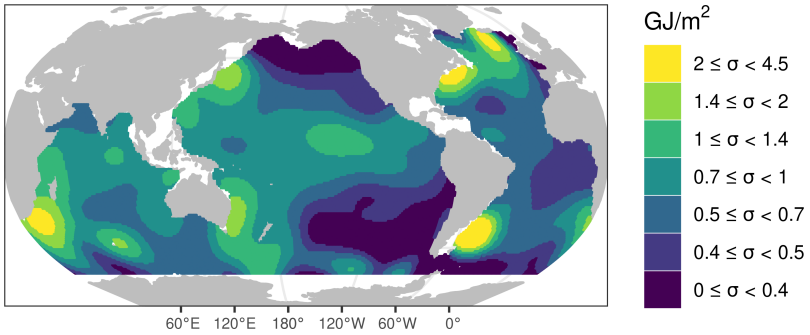
(a) Effective longitudinal range γ_{lon}



(b) Effective latitudinal range γ_{lat}



(c) Process standard deviation $\sqrt{\phi}$



(d) Nugget standard deviation σ

Figure 2.7: Initial parameter fields obtained from fitting smooth hyperparameter surfaces to the parameter fields obtained from a moving-window approach.

	Effective Range	$q_{.25}$	$q_{.5}$	$q_{.75}$	Units
γ_{lat}	39.97°	0.89	2.50	6.98	Degrees
γ_{lon}	44.93°	0.88	5.07	29.08	Degrees
σ^2/ϕ	36.59°	.0034	0.04	0.47	Unitless
$\sqrt{\phi}$	39.24°	1.29	2.25	3.92	GJ/m^2
μ_{2007}	34.88°	19.70	49.24	78.78	GJ/m^2
β	34.88°	-1.53	0.00	1.53	(GJ/m^2)/year

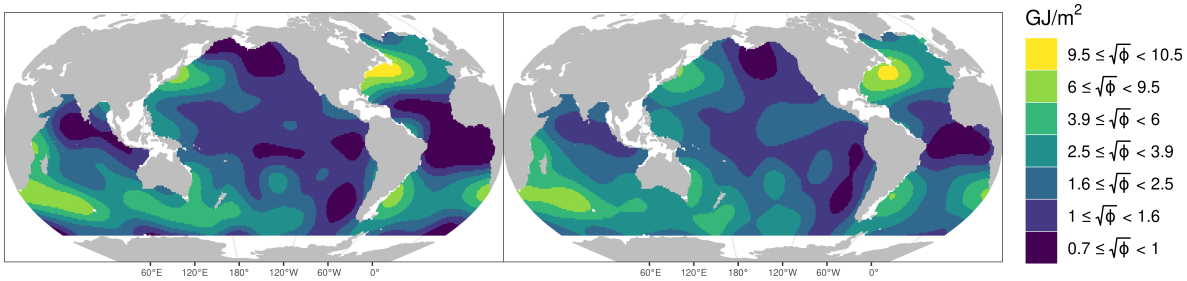
Table 2.2: Basis field hyperparameters as estimated using maximum likelihood estimation on parameter fields obtained from a $20° \times 20°$ moving window estimation. The range hyperparameter and the correlation length scale fields γ_{lat} and γ_{lon} are reported in terms of effective range, which is the distance at which the correlation is 0.05. Quantiles are reported due to the non-symmetric nature of the log-normal distribution used for the covariance parameters. The mean of the trend parameter field β was manually constrained to be zero to avoid influencing its estimation in the Bayesian procedure.

over a dense grid for display purposes. Similar to Table 2.2, standard deviations $\sqrt{\phi}$ and effective ranges γ_{lon} and γ_{lat} are shown rather than the actual parameter fields. While this configuration should not be used to quantify global OHC due to the fact that the estimates were derived assuming local stationarity, comparisons to the 2016 anomaly field displayed in Figure 2.2 yield insights regarding the ocean heat content correlation structure. First of all, it can be seen from the initial configuration that the correlation length scales around the equatorial regions are much larger than at higher latitudes, and are particularly large in the equatorial Pacific. This phenomenon can be seen in the heat content anomaly field for 2016 through the coherent structures of positive anomalies particularly visible in the eastern equatorial Pacific. These anomalies primarily extend in the longitudinal direction and to a smaller extent in the latitudinal direction. Anisotropy is reflected in the initial configuration, where it can be seen that the scale of the effective longitudinal correlation lengths is

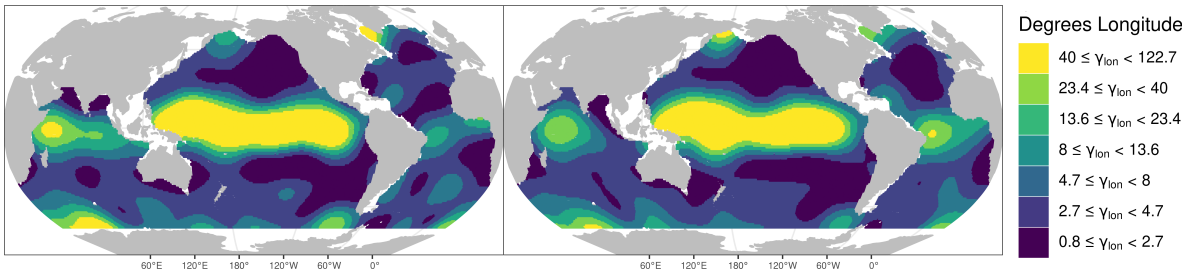
generally much larger than those in the latitudinal plot, a difference that is particularly stark in the equatorial Pacific. The phenomenon of latitudinal correlation lengths being generally smaller than longitudinal correlation lengths makes sense from a physical perspective, as changes in latitude are associated with changes in climate to a much greater degree than changes in longitude.

The areas of the ocean that appear to have particularly high variability in the 2016 anomaly field roughly correspond to the areas of high variance in the initial configuration. Four regions in particular stand out in both of these plots; the North Atlantic east of the United States, the Pacific Ocean east of Japan, the South Atlantic east of Argentina, and a large swath of ocean moving towards the south-east from South Africa to South Australia. It should be noted that most of these regions align with known ocean currents, in particular the Gulf Stream of the North Atlantic, the Kuroshio Current off the coast of Japan, the South Atlantic current off of Argentina, and the Arctic circumpolar current in the Southern Ocean [Colling, 2001]. This is unsurprising, as the currents generate large anomalies of either positive or negative sign depending on the particular position and behavior of the current at a given time.

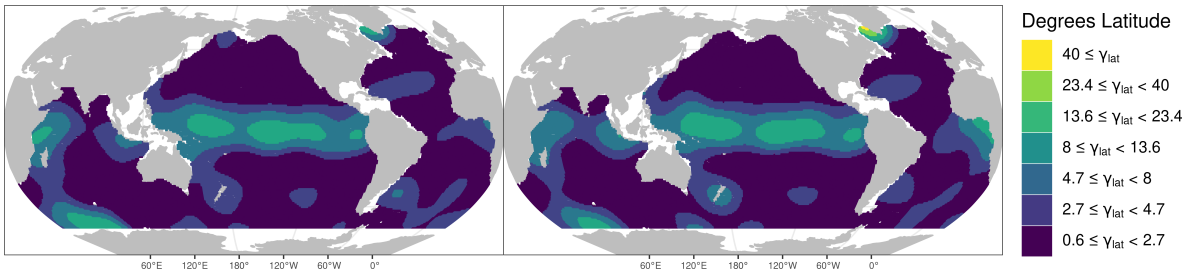
In theory, the choice of initial configuration will not impact the results of the MCMC sampler if a sufficiently long chain is sampled. However, running the sampler on random initial configurations did not converge after 10,000 iterations. Starting from this initial configuration considerably improves the sampler's convergence properties as will be discussed in Section 2.6.2. Sampling the parameter fields and the hyperparameters simultaneously is particularly challenging due to the complexity of the posterior distribution. To avoid this issue, the hyperparameter values obtained from the moving window approach were treated as fixed and not sampled. This can be interpreted as imposing an informative prior which constrains the variability of the parameter fields to reasonable levels.



(a) $\sqrt{\phi}$ 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile samples



(b) γ_{lon} 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile samples



(c) γ_{lat} 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile samples

Figure 2.8: Samples from the posterior distribution with mean values corresponding to the fifth (left figures) and ninety-fifth (right figures) percentiles of the spatial averages of each field from the posterior distribution.

2.6.2 Posterior Distributions

Starting from the initial configuration obtained in the previous section, the MCMC sampler was run for 20,000 iterations on a desktop computer with $16GB$ of RAM running Ubuntu 20.04 and using ten cores for the parallel computation of the likelihood evaluations and OHC estimations. The ocean heat content distribution was computed using Equation (2.13) on a $1^\circ \times 1^\circ$ grid every ten iterations. To assess convergence, the Heidelberger and Welch test [Heidelberger and Welch, 1981] was run on the sequence of log posterior densities, yielding convergence after a burn-in period of 5,401 iterations. The results that follow are computed on the posterior samples with the burn-in period removed.

To get a sense for the posterior distribution of the parameter fields, for each parameter field the average value across space was computed for each sample, and fields corresponding to the 5^{th} and 95^{th} percentiles of the distribution were identified. These parameter configurations for the standard deviation $\sqrt{\phi}$, effective latitudinal range θ_{lat}, and effective longitudinal range θ_{lon} are displayed in Figure 2.8. Note that these are not the point-wise 5th and 95th percentiles of the parameter estimates, which would not retain the spatial structure present in each sample, but rather spatially coherent members of the posterior distribution. The most noticeable difference between these samples and the initial configuration displayed in Figure 2.7 are the higher longitudinal range parameters in the samples, particularly noticeable in the equatorial West Pacific. This could be due to the fact that the moving-window parameter method may underestimate high values of the range parameter due to the fact that its estimates are based on data from only a $20^\circ \times 20^\circ$ window. Otherwise, the samples appear to be largely similar to the initial parameter fields, further highlighting the benefit of starting the MCMC sampler at a plausible initial configuration.

While the 5^{th} and 95^{th} percentile configurations in Figure 2.8 are fairly similar to each other for each of the parameter fields, the subtle differences between the percentile maps show that there are regions with higher posterior variability. In the process standard deviation

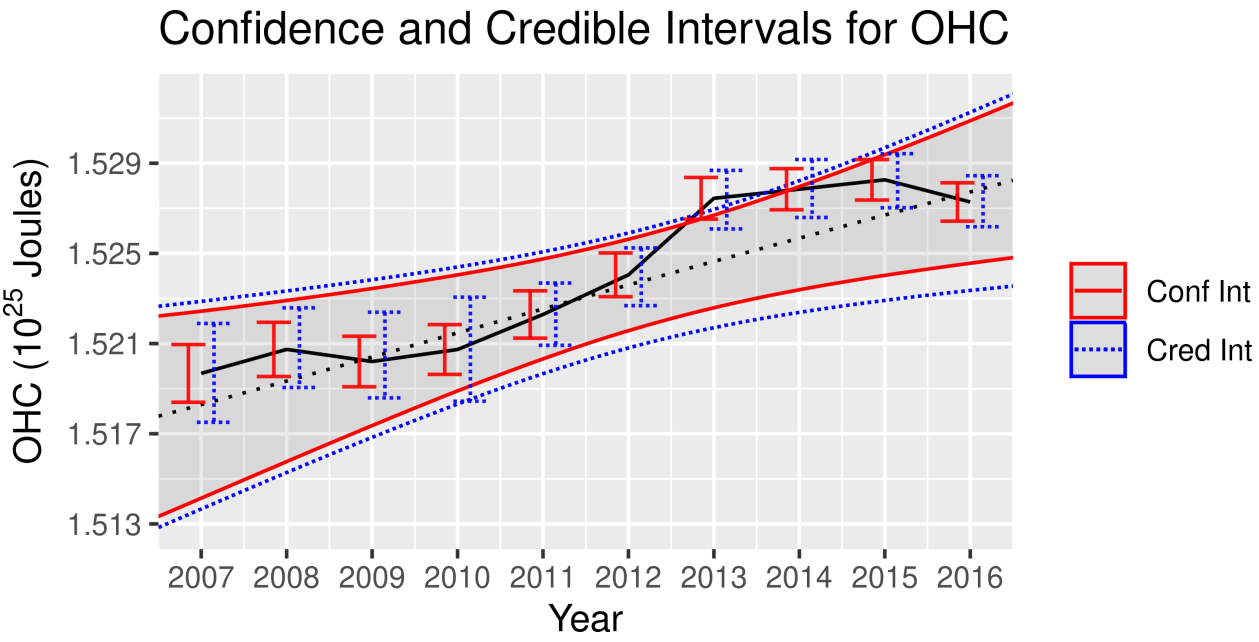


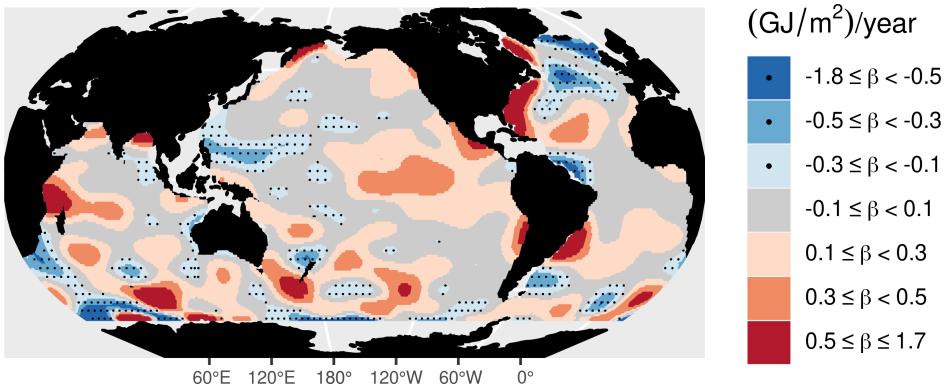
Figure 2.9: Estimation of uncertainty in the ocean heat content values for January of each year. Confidence intervals in solid red correspond to uncertainty arising from the variability of the heat content field. Credible intervals in dotted blue include the additional uncertainty incurred from parameter estimation. Intervals and posterior mean are also shown for the increasing trend.

maps in Figure 2.8a, differences can be seen in the value and areal extent of the several highly-variable current regions discussed before, indicating some uncertainty in the marginal standard deviation values in these areas. For the longitudinal range maps in Figure 2.8b, variability can be seen in the higher-ranged area in the western Pacific, which is the region that changes the most between the initial configuration and the posterior samples. It can also be seen that variability in the distribution of longitudinal ranges in the Southern Ocean, which is likely due to the fact that parameters in this region are more difficult to estimate due to the smaller number of observations near the extent of Antarctic sea ice, as well as the large variability in the data. For the posterior maps for latitudinal range in Figure 2.8c, it can be seen that variability in the value and extent of the high-ranged area in the equatorial regions, as well as in smaller areas such as south-east of Argentina and around New Zealand. Overall, however, the variability in the posterior distribution appears to be small for each of the parameter fields, indicating that the uncertainty in the ocean heat content and trend estimations induced by uncertainty in the estimation of the parameters will be relatively low.

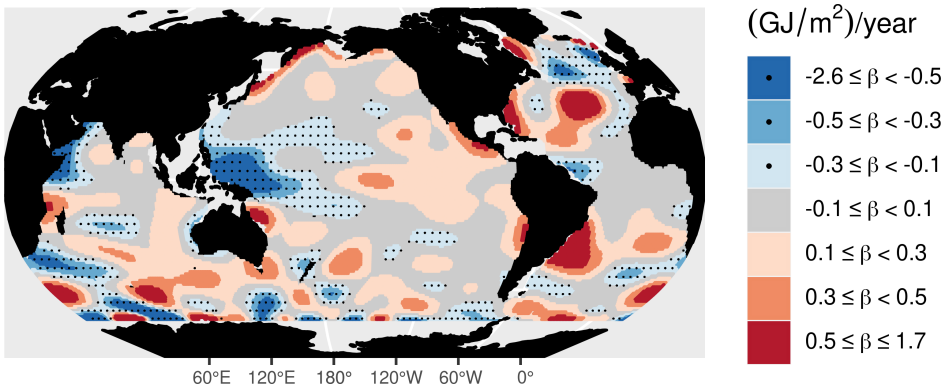
The ocean heat content mean and variance were calculated for each year for every ten samples after the burn-in period. For each year, the median OHC value over the posterior samples and the 95% confidence interval $\mu^{\text{median}}_{\text{OHC}_{yr}} \pm 1.96\sigma^{\text{median}}_{\text{OHC}_{yr}}$ were calculated; these are shown as solid red bars in Figure 2.9. This represents the uncertainty range induced by the interpolation of the observations to unobserved locations and can be interpreted as a frequentist confidence interval as it does not account for the priors on the parameters. To compute the Bayesian credible interval for each year, for each iteration t of the sampler 100 heat content values were re-sampled from the marginal OHC distribution $\mathcal{N}(\mu^{(t)}_{\text{OHC}_{yr}}, \sigma^{(t)}_{\text{OHC}_{yr}})$. Then 5^{th} and 95^{th} percentiles were computed from the pooled re-resampled OHC values. The re-sampling step ensures that the credible intervals accurately take into account both the interpolation uncertainty from $\sigma^{(t)}_{\text{OHC}_{yr}}$ as well as the parameter uncertainty from the variability of $\mu^{(t)}_{\text{OHC}_{yr}}$ over the samples. These intervals are displayed as dotted blue bars

in Figure 2.9. It can be seen that both of the intervals are wider in earlier years, which is to be expected since there are fewer data points in those years. The credible intervals are meaningfully wider than the confidence intervals in earlier years, although the difference becomes small in later years, which is to be expected since the parameter values have a greater effect on the interpolation when there are larger gaps in the data. An analogous procedure was done to obtain uncertainty intervals for the estimated trend; for this parameter, the entire trend field was re-sampled using the posterior covariance matrix conditioned on the other parameters at each iteration. The 95% credible interval was obtained by identifying the 2.5% and 97.5% percentiles of the integrated trend values over the re-sampled fields. The 95% confidence interval, computed with the configuration corresponding to the median integrated trend value, is displayed with solid red lines in Figure 2.9, while the 95% credible interval is displayed as dotted blue lines. The globally integrated trend is positive and highly significant, with a p-value of .00165 obtained from the credible interval.

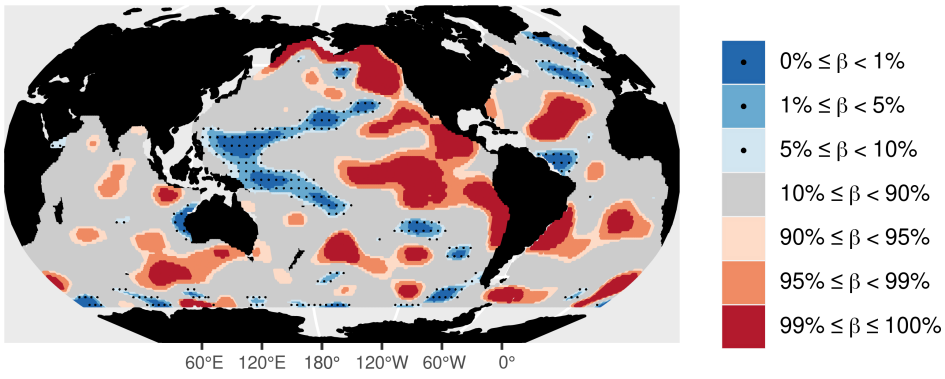
Re-sampling not just the globally integrated trend values but the entire field yields additional information about the spatial variability in the ocean heat content trend. Figure 2.10 displays samples of the trend fields whose globally integrated values equal the 5th (panel a) and 95th (panel b) percentiles of the re-sampled distribution. It can be seen that variability in the trend field is higher than that in the correlation and variance parameters in Figure 2.8. Notable features include the cooling trend in the western Pacific and the warming trend in the eastern Pacific, although there appears to be variability in the intensity and areal extent of these trends. There is more confidence in a warming trend in the ocean south-east of Brazil with little apparent variation in the intensity or spatial extent between the two percentile maps. The North Atlantic, on the other hand, displays variable spatial patterns, with the 5th percentile field showing a negative trend southeast of the United States that is not present in the 95th percentile map, and while the two percentile fields agree on a cooling trend in the far North Atlantic they differ on its intensity. Other regions where the trend pattern appears largely uncertain are in the Indian Ocean and the Southern Ocean.



(a) 5^{th} percentile trend field



(b) 95^{th} percentile trend field



(c) Posterior probability of positive trend

Figure 2.10: Sub-figures (a) and (b) show samples of the trend posterior whose integrated values are the fifth and ninety-fifth values of the posterior distribution, which are 4.660×10^{21}J/year and 16.032×10^{21}J/year respectively. Sub-figure (c) gives the posterior probability that an ocean pixel is gaining heat; values less than 10% indicate confidence that the trend is negative.

These two percentile maps can only give an approximate view of the variability in the trend field, as there is a high-dimensional space of fields whose globally integrated values lie at the 5th and 95th percentiles of the posterior distribution. To obtain a better idea of agreement in the posterior, the percentage of the re-sampled trend fields that agree on the sign of the trend for each grid-cell of a $1^\circ \times 1^\circ$ grid was calculated. This is equivalent to the posterior probability of a positive trend value, or $P(\beta(x) > 0|\text{observations})$, and the results are displayed in Figure 2.10c. It can be seem that there is agreement on certain spatial features such as the western Pacific cooling trend and the eastern Pacific warming trend. As expected, there is high-level agreement on the warming trend in the ocean south-east of Brazil, and while there is agreement on a positive trend region in the mid-latitude North Atlantic and on the cooling trend in the high-latitude North Atlantic there is no agreement on the trend in the area of the ocean south-east of the United States. There is not any strong agreement on general trends in the Indian and Southern Oceans, however, there are isolated bubbles of confident warming and cooling scattered throughout these regions. It would be expected that these small-scale features would no longer be significant if more years of observations were used to fit the model, which is reserved for Chapter 3.

2.7 Model Validation

Leave-one-float-out (LOFO) cross-validation is used to validate the model. This metric, for each float, removes the observations corresponding to that float and then uses the remaining locations to predict the values at the removed locations. The motivation for LOFO validation is that, in the Argo data, the distance between the three profiles observed by a single float during a month is generally less than that between observations from different floats. Removing the entire float's worth of observations allows the cross-validation results to evaluate how well the model is doing at estimating heat content at unobserved areas of the ocean. One downside to this approach is that the spatial area from which the floats are removed

will be variable due to the fact that floats are not equally distributed around the ocean. The cross-validation performance was evaluated using mean absolute error (MAE), root-mean-squared error (RMSE), and continuous ranked probability scores (CRPS) [Gneiting et al., 2005]. While the MAE and RMSE metrics assess the accuracy of the predictions, the CRPS metric evaluates precision as well by penalizing higher standard errors in the predictions.

These cross-validation methods were evaluated using the maximum posterior density configuration from the global sampler, and Figure 2.11 shows the LOFO CRPS scores averaged over observations from all years within each cell of a $5^\circ \times 5^\circ$ grid. The globally-averaged LOFO scores for the full model are displayed in the first row of Table 2.3. A windowed cross-validation approach as described in Section 2.8 was also performed, with results showing patterns similar to the LOFO results displayed in Figure 2.12 and Table 2.4. It can be seen from Figure 2.11 that while the validation errors are low in the equatorial regions where the variance is low and the spatial correlation is high, the model performs worse in the high variance and low spatially correlated areas corresponding to the currents.

Due to the variability of these regions, there is a limit to the predictive ability of any model, and as such the raw validation scores are difficult to put into context without a reference model for comparison. To compare the proposed model to a commonly cited approach in the climate community the non-stochastic approach of Levitus et al. [2012] was implemented to compute predictions and associated uncertainties in the cross-validation procedures. The Levitus et al. [2012] approach was also used to compare predicted ocean heat content values and trends, with results presented in Figure 2.12 and Table 2.4 in Section 2.8. As can see in Table 2.3, the fully non-stationary model produces a 11.2% improvement over the Levitus et al. [2012] approach in terms of RMSE in the LOFO validation. Following the logic of Kuusela and Stein [2018], assuming a $1/\sqrt{n_{\text{obs}}}$ rate of convergence for the Gaussian process predictions, these percent improvements can be seen as equivalent to a 24% increase in the number of observations, or an addition of around 650 floats.

While it is apparent that the fully non-stationary cylindrical model outperforms the

	MAE	RMSE	CRPS
Fully Non-stationary	1.610 (9.34%)	2.708 (11.19%)	1.330 (8.15%)
Stationary σ^2/ϕ	1.612 (9.24%)	2.699 (11.47%)	1.324 (8.56%)
Non-stationary Isotropic	1.625 (8.50%)	2.731 (10.43%)	1.344 (7.22%)
Stationary θ_{lat}	1.633 (8.04%)	2.738 (10.20%)	1.347 (7.03%)
Stationary θ_{lon}	1.655 (6.81%)	2.762 (9.42%)	1.364 (5.79%)
Stationary ϕ	1.667 (6.12%)	2.701 (11.43%)	1.364 (5.85%)
Stationary Spatiotemporal	1.705 (3.98%)	2.850 (6.53%)	1.603 (-10.66%)
Stationary Isotropic	1.718 (3.26%)	2.865 (6.03%)	1.618 (-11.74%)
Fully Stationary	1.742 (1.94%)	2.758 (9.54%)	1.431 (1.19%)
Levitus et al.	1.776 (0.00%)	3.049 (0.00%)	1.448 (0.00%)

Table 2.3: Leave-one-float-out cross-validation scores for the fully non-stationary anisotropic model, restrictions to stationarity in each of the parameter fields, isotropic models, and a spatio-temporal model. Percentage improvements over the reference model [Levitus et al., 2012] are displayed in parentheses. Units are GJ/m^2.

reference model in cross-validation, which non-stationary aspects of the proposed model contribute the most to the improvement is also of interest. To assess this, for each of the correlation and variance parameter fields, the model was re-fit under the constraint that the parameter field is spatially constant. These models were fit by running the **BayesianOHC** sampler with the stationarity constraint imposed. Initial configurations were obtained by adjusting the last sample of the full MCMC chain so that the newly stationary field has the median value of the non-stationary field.

The cross-validation scores with relative improvements over the reference model are shown in Table 2.3. In terms of mean absolute error, the constrained models have improved cross-validation scores over the fully non-stationary model, indicating the importance of accounting for spatial structure in the model parameters for the ocean heat content field. However,

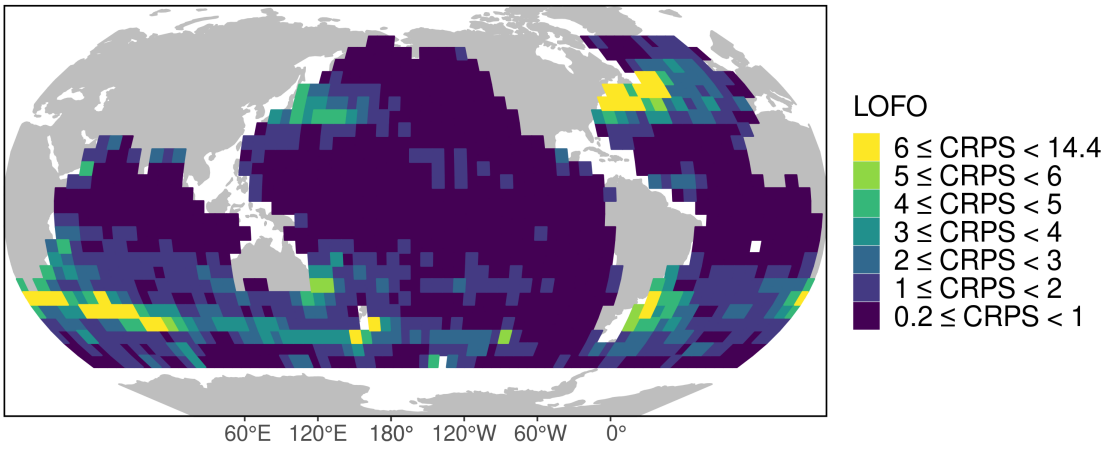


Figure 2.11: Average LOFO CRPS over all years for each grid-cell of a $5° \times 5°$ grid.

the stationary inverse signal-to-noise and marginal variance models perform slightly better than the full model in terms of RMSE scores, with the stationary signal-to-noise model also having a slightly improved CRPS score. This is likely due to the fact that the fully non-stationary model has higher average values of these parameters than the corresponding stationary models, with the stationary signal-to-noise model having a σ^2/ϕ value of .18 versus the average full model value of .24, and the stationary variance model having a $\sqrt{\phi}$ value of 2.28 versus the average full model value of 2.41. Unlike MAE, the RMSE and CRPS values take into account error variance and prediction uncertainty, which can explain why the stationary models perform slightly better with regard to those metrics. It should be noted however that non-stationary modeling of the variance parameters is still desirable, as keeping those parameters stationary will likely underestimate the uncertainty in highly variable regions. Increases in all three cross-validation metrics are observed when θ_{lat} and then θ_{lon} are forced to be stationary, which is as expected given that these parameters, especially θ_{lon}, vary substantially over the domain (Figure 2.8).

The proposed approach involves the use of a cylindrical approximation for the sphere in order to allow for separate correlation length scales in the latitudinal and longitudinal

directions, which is important given the anisotropic behavior of ocean heat content. It is common in the statistics literature to instead treat the sphere as embedded in $\mathbb{R}^3$ space and use the euclidean chord-length distance within an isotropic framework. There are several existing packages including **GpGp** [Guinness and Katzfuss, 2018], **GPVecchia** [Katzfuss et al., 2020c], and **BayesNSGP** [Turek and Risser, 2019] that allow for isotropic Gaussian process modeling using this metric. While there is not a natural way to represent anisotropy using correlation length-scales with $\mathbb{R}^3$ coordinates, it is worthwhile to examine how well these approaches perform in comparison to ours at cross-validation. To compare against a stationary model the "fit_model" function from the **GpGp** package was used with the "matern_sphere" covariance option to fit an isotropic Matérn model to the data in each year independently. As this package does not immediately allow for a Gaussian process mean-field as used in the proposed model, a mean-field parameterization consisting of linear and quadratic terms of latitude and longitude was used instead. The results of this model fit are displayed in the "Stationary Isotropic" row of Table 2.3. While this model has an improved MAE score over the fully stationary anisotropic cylindrical fit, it performs worse in all metrics than the non-stationary models fit using **BayesianOHC**. As Guinness [2021] have noted that isotropic spatio-temporal modeling yields a higher likelihood than non-stationary spatial-only models, the **GpGp** package was used to fit a stationary spatio-temporal model to the data, shown in the "Stationary Spatiotemporal" row of Table 2.3. While this performs better than a spatial-only isotropic model, it still underperforms the non-stationary cylindrical models.

To evaluate the difference between the isotropic and anisotropic approaches in the non-stationary setting, the **BayesianOHC** package was used to fit a non-stationary isotropic model using the chord-length distance to the full set of Argo data, with the LOFO cross-validation results displayed in the third row of Table 2.3. It can be seen that while the non-stationary isotropic model outperforms most of the other models, the full model achieves superior cross-validation scores. As the only difference between these two models is the cylindrical anisotropy component, this improvement shows the value of representing latitude-

longitude anisotropy when modeling ocean heat content.

2.8 Windowed Cross-Validation Results

In addition to the leave-one-float-out (LOFO) cross-validation study in Table 8, cross-validation was performed on the same set of models using a windowed approach, where for each observation the observed locations within a $2^\circ \times 2^\circ$ window around the location are withheld in order to get a sense for how well the models predict at unobserved locations. The continuous ranked probability scores (CPRS) averaged over each $5^\circ \times 5^\circ$ grid-box are shown in Figure 2.12. Similar patterns are seen in the analogous Figure 7 for LOFO cross-validation although the LOFO MAE scores are 10.9% larger on average.

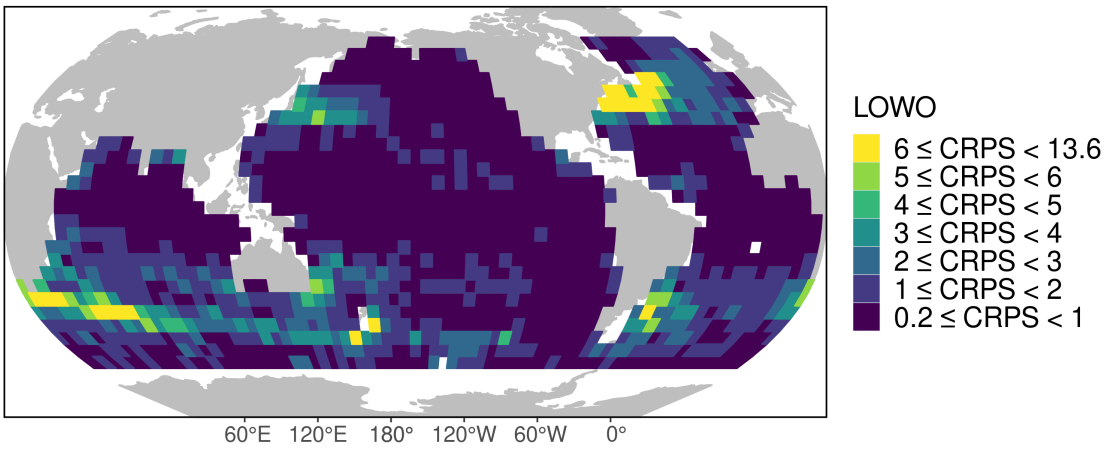


Figure 2.12: Average windowed cross-validation CRPS values over all years for each grid-cell of a $5^\circ \times 5^\circ$ grid.

The windowed cross-validation mean absolute error (MAE), root-mean-squared error (RMSE), and continuous ranked probability scores (CRPS) are displayed in Table 2.4 along with percent improvements over the Levitus et al. [2012] reference model. Trends similar to Table 2.3 can be seen, with the reference model generally under-performing the other models

and the fully non-stationary anisotropic model generally over-performing the other models.

	MAE	RMSE	CRPS
Fully Non-stationary	1.451 (11.11%)	2.375 (12.09%)	1.174 (12.13%)
Stationary σ^2/ϕ	1.458 (10.63%)	2.357 (12.76%)	1.175 (12.03%)
Stationary θ_{lat}	1.471 (9.83%)	2.403 (11.05%)	1.188 (11.03%)
Nonstationary Isotropic	1.475 (9.62%)	2.404 (11.02%)	1.195 (10.54%)
Stationary θ_{lon}	1.499 (8.13%)	2.427 (10.18%)	1.212 (9.29%)
Stationary Spatiotemporal	1.534 (6.01%)	2.469 (8.60%)	1.432 (-7.19%)
Stationary ϕ	1.541 (5.56%)	2.446 (9.48%)	1.246 (6.72%)
Stationary Isotropic	1.544 (5.41%)	2.477 (8.31%)	1.445 (-8.15%)
Fully Stationary	1.610 (1.35%)	2.479 (8.26%)	1.298 (2.80%)
Levitus et al.	1.632 (0.00%)	2.702 (0.00%)	1.336 (0.00%)

Table 2.4: Windowed cross-validation scores for the fully non-stationary cylindrical distance model, restrictions to stationarity in each of the parameter fields, models using the chord-length distance, and a stationary spatio-temporal model. Chordal distance models and the spatio-temporal model are isotropic.

2.9 Comparison with Results Obtained from the Method Described in Levitus et al.

The approach to interpolating ocean heat content and calculating standard errors used in Levitus et al. [2012] is commonly cited in the climate community. In this method, the predicted ocean heat content value at an unobserved point is the weighted sum of observations within a fixed radius around the point. The weights are determined by an exponential function of the negative squared distance between the locations. Standard errors are calculated by computing the weighted standard deviation of observations within the radius with the

assumption that the errors at different locations are independent.

While Table 2.3 and Table 2.4 show that the fully non-stationary Gaussian process cylindrical distance model out-performs the Levitus et al. method at cross-validation, it is of interest to compare the predicted ocean heat content estimates and uncertainty values between the two methods. Figure 2.13 displays the predicted global OHC values and 95% uncertainty intervals obtained using the Levitus et al. [2012] method along with the posterior medians and credible intervals obtained using the **BayesianOHC** fully non-stationary cylindrical distance model. After using the Levitus approach to compute global heat content values for each year, the implied trend was estimated using ordinary least-squares and the corresponding confidence interval was calculated. The resulting estimated trends and uncertainty intervals are displayed in Table 2.5.

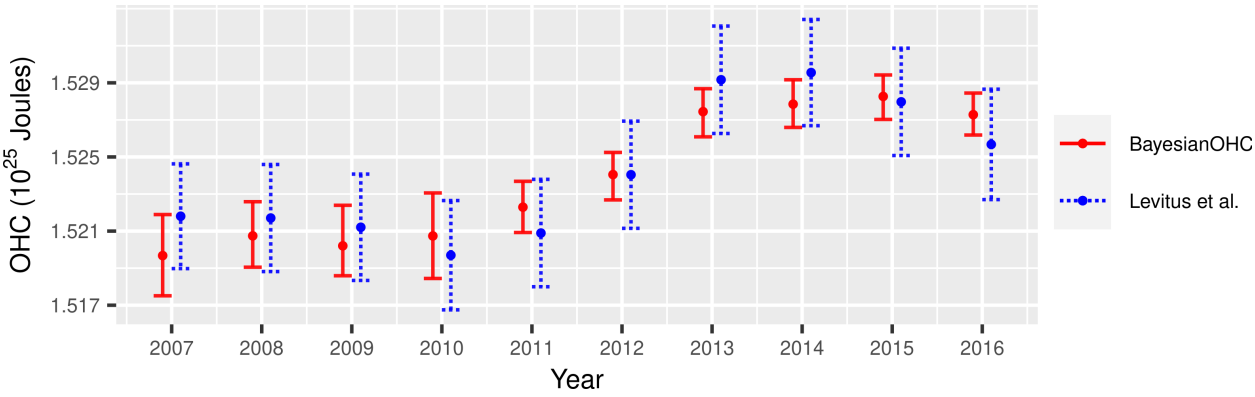


Figure 2.13: Comparison of predicted ocean heat content values and uncertainty intervals between the non-stationary cylindrical method and the Levitus et al. method.

It can be seen that both methods are fairly similar in terms of predicting yearly heat content values, however, the standard errors using the Levitus approach seen in Figure 2.13 are substantially larger than those obtained from **BayesianOHC**. This is to be expected as unlike in a Gaussian process framework, the Levitus et al. [2012] approach does not take into consideration correlations between locations and rather assumes that the errors in distinct gird-boxes are independent. On the other hand, it can be seen in Table 2.5 that

the **BayesianOHC** credible interval is wider than the ordinary least-squares confidence interval. This is also not surprising due to fact that the uncertainty from fitting ordinary least-squares to the Levitus predictions does not take into account the uncertainty induced from the spatial interpolation, and as a result the confidence interval underestimates the uncertainty in the trend. This is in contrast to the **BayesianOHC** credible intervals which are estimated within the context of a hierarchical model, and as such take into account the uncertainty induced by the interpolation of the heat content field as well as the additional uncertainty from estimating the other parameters in the model.

Despite the wider uncertainty range in the **BayesianOHC** estimated trend, the larger median trend estimate yields a smaller p-value than the Levitus et al. [2012] least-squares fit. If hypothesis testing was used, the Levitus et al. [2012] approach would fail to reject a null hypothesis of no trend at a one percent significance threshold whereas the **BayesianOHC** method would conclude that there is sufficient evidence that the oceans have gained heat between 2007 and 2016. This demonstrates the value of using statistical modeling to quantify ocean heat content as the non-stationary cylindrical distance model more accurately quantifies heat content and its variability with a meaningful impact on the significance of the trend.

	2.5%	Median	97.5%	Width	p-value
Levitus et al.	2.753×10^{21}	9.215×10^{21}	15.677×10^{21}	12.924×10^{21}	0.01105
BayesianOHC	3.537×10^{21}	10.321×10^{21}	17.139×10^{21}	13.601×10^{21}	0.00165

Table 2.5: Estimated trend and associated 95% confidence interval for the ordinary least-squares trend fit to the heat content predictions computed using the Levitus et al. [2012] method and the median posterior trend and associated 95% credible interval using the fully non-stationary **BayesianOHC** fit. All units are in J/year.

2.10 Conclusion

The kernel-convolution hierarchical Bayesian Gaussian process model developed in this chapter is able to provide uncertainty quantification in ocean heat content values, as well as in the heat content trend, that could not be produced using previous methods. The Gaussian process approach allows us to quantify the uncertainty induced from having to interpolate the ocean heat content field over the domain using information from limited observations. The kernel-convolution covariance structure allows us to capture the non-stationary covariance properties of the ocean heat content field through spatially-varying parameter fields. These convolutions are made possible by using the cylindrical distance, an acceptable distance metric for this application because the model does not extend to the poles and the longitudinal correlation length scales can vary with the sphere's changing radius over latitude. The cylindrical distance metric is a novel approach to representing anisotropy on the sphere that is not possible using an $\mathbb{R}^3$ representation. The resulting fitted covariance model has been shown to achieve validation scores superior to simpler methods, including a fully non-stationary isotropic model, which demonstrates the value of using the cylindrical distance approach to model anisotropy. The hierarchical Bayesian approach yields posterior distributions for each of the parameter fields which allows us to assess the uncertainty in their estimation. This additional source of uncertainty is then represented in the posterior distribution of ocean heat content and its trend. The model's non-stationary representation of the mean and trend fields allows for estimating the spatial distribution of certainty regarding local warming or cooling trends. The models have been fit in a computationally efficient way using the Vecchia process, which has been shown to yield a sufficiently accurate approximation to the full Gaussian process. Functionality for analyzing ocean heat content has been developed as the R package **BayesianOHC** along with scripts for reproducing the results presented in this paper. While the **BayesianOHC** package has many utilities that are specifically useful for modeling Argo data, many of the model-fitting functions are general and can be used for fitting non-stationary Bayesian models for a variety of applications.

The results presented in this Chapter are limited in the amount of data considered, specifically to January over the years 2007-2016 and for floats whose profiles extend to 1,900m of depth. The restriction to January was done in order to focus on the spatial components of the framework. The extension of the framework to the spatio-temporal setting and to data in all months will be developed in the next chapter.

CHAPTER 3

Spatio-Temporal and Seasonal Modeling of the Ocean Heat Content Field

3.1 Introduction

The previous chapter has introduced a hierarchical Bayesian Gaussian process framework for modeling the spatially non-stationary correlation structure of the ocean heat content field. This model was fit to a subset of the data such that seasonal variation in the mean-field and temporal correlations in the anomalies could be ignored, producing a posterior distribution for the year-to-year trend that takes into account sampling uncertainty, uncertainty from variability in the process, and uncertainty from parameter estimation. This chapter extends the spatial-only model to the spatio-temporal setting in order to take advantage of the information presented by spatial correlation in the data as well as the additional data from incorporating all months. The treatment begins with updating the notation for representing the data in the spatio-temporal setting in Section 3.2. Section 3.3 introduces two improvements to the spatial-only model to account for temporal properties in the mean and anomaly fields. The first development in Section 3.3.2 is of a mean-field that models smooth seasonal variation simultaneously with spatial variation through a spatio-temporal Gaussian process prior with a "circular" correlation structure over time. Section 3.3.1 introduces a temporal dimension to the kernel-convolution process with spatially-varying temporal correlation length scale parameters analogous to the latitudinal and longitudinal convolutions presented in Chapter 2.

This chapter proceeds to apply the spatio-temporal framework to Argo data from all months over the years 2007-2021. Parameter fields for initializing the MCMC sampler are derived using an updated moving window approach in Section 3.4. In Section 3.5 the spatial Vecchia process used in Chapter 2 as a computationally efficient approximation for the full Gaussian process is augmented to account for the spatio-temporal structure of the data. This requires a re-evaluation of the size of the conditioning sets needed to obtain a sufficiently accurate representation, which is done through a study evaluating the accuracy of the posterior distributions of the mean and trend fields for different conditioning set sizes. Finally,

results from the posterior distribution are presented and discussed in Section 3.6, including the posterior correlation parameter fields, the spatially and seasonally varying mean-field, distributions for the ocean heat content values at each month, and the spatially-varying trend field. The resulting estimates for the trend in ocean heat content are consistent with the results in the previous chapter but with a much higher level of confidence due to the increase in data and the information provided by the temporal structure.

3.2 Spatio-Temporal Data

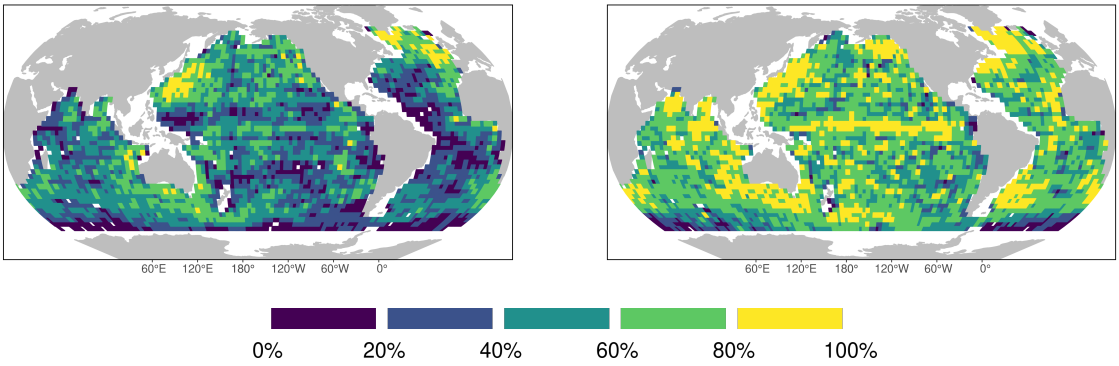
To extend the framework to data from all months, the notation must be adjusted to represent each location as a spatio-temporal point and to re-define heat content as a field varying continuously over time. Let $\mathbf{x} = (x_{\text{lat}}, x_{\text{lon}}, x_{\text{time}}) \in \mathbb{CL} \times \mathbb{R}$, where as before $\mathbb{CL} \equiv \mathbb{R} \times \mathbb{S}$ represents latitude and longitude in Euclidean and spherical coordinates respectively, and x_{time} represents time in days. Furthermore, let $z \in \mathbb{R}^+$ denote depth from the surface in meters. Define the three-dimensional temperature field at spatio-temporal location $\mathbf{x}$ and depth z as $T(\mathbf{x}, z)$ as measured in degrees Celsius. Similarly to Chapter 2, the spatio-temporal field is defined as heat content values integrated over depth as calculated using following formula:

$$H(\mathbf{x}) = \int_{0\text{m}}^{2000\text{m}} \rho c_\rho T(\mathbf{x}, z)\, dz. \tag{3.1}$$

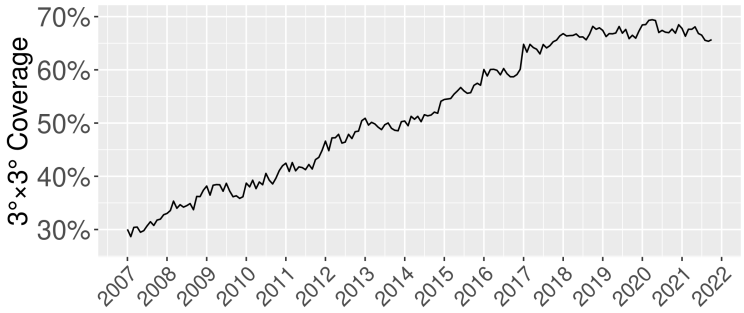
Here, ρ is seawater density in kg/m^3 and c_ρ is the specific heat of seawater in J/(kg$^\circ$ C) [Dijkstra, 2008]. The units of $H(\mathbf{x})$ are J/m^2. The total global ocean heat content or OHC is a spatially-integrated quantity that varies continuously over time:

$$\text{OHC}(t) = \int_{[x_{\text{lat}}, x_{\text{lon}}] \in S_{\text{mask}}} H([x_{\text{lat}}, x_{\text{lon}}, t])\, d\mathbf{x} \tag{3.2}$$

where S_{mask} is the ocean's surface masked following Roemmich and Gilson [2009] and displayed in Figure 2.1. Again, floats with maximum depths that are not within 100m of 2000m are excluded to avoid having to interpolate the heat content field over depth.



(a) The percentage of months containing at least one observation within each $3° \times 3°$ grid-cell. Left panel is averaged over the years between 2007 and 2014, and the right panel is averaged over the years 2015 to 2021.



(b) The average number of $3° \times 3°$ grid-cells with at least one observation displayed over months.

Figure 3.1: Spatial distribution of data coverage at the $3° \times 3°$ grid-cell resolution over 2007 to 2014 and 2015 to 2021 (a), and the increase in data coverage over time (b).

The data under consideration here range from January 2007 to November 2021, and after applying the mask and depth filters there are 1,111,023 total ocean heat content observations. These observations are not uniformly distributed over space, with some regions frequently seeing more floats than others, nor over time, with an increase in the number of floats deployed over time resulting in more recent years having more observations than earlier years. To examine the spatial distribution of observations and how data coverage has increased over recent years, the spatio-temporal domain was partitioned into $3^\circ \times 3^\circ$ grid-cells for each month within the time range. For each grid-cell-month pair, an indicator function was applied for whether or not it contained at least one observation; this is defined as the $3^\circ \times 3^\circ$ resolution "data coverage". To visualize the spatial distribution of the data coverage, the data was partitioned between subsets occurring between 2007 and 2014 and observations occurring between 2015 to 2021. For each of these subsets, the data coverage averaged over all months was calculated for each grid-cell. The resulting spatial distributions of data coverage are displayed in Figure 3.1a, with data coverage from 2014 and earlier on the left panel and 2015 and later on the right. It can be seen that in each of the time periods, data coverage is higher in certain areas than others, with the Pacific Ocean east of Japan and the North Atlantic exhibiting high data coverage over each time period. While these regions have over 80% data coverage in both time periods, grid-cells with less than 20% data coverage can be observed in most regions in the earlier time period. In the later time period, grid-cells with less than 20% data coverage occur rarely and most regions have greater than 60% data coverage. A notable exception is the far Southern Ocean, which exhibits low levels of data coverage even in the more recent time frame.

To get a further sense of the increase in data coverage over time, coverage was averaged spatially for each month in the full time period with results displayed in Figure 3.1b. It can be seen that data coverage increases continuously until about 2018, when it begins to stabilize at values nearing 70%. Coverage ranges from 30% in the first month of January 2007 to 65.67% to the last month of November 2021. In the dataset under consideration

here, the number of unique floats with valid ocean heat content profiles ranges from 1,061 in January 2007 to 3,164 in November 2021.

3.3 Model Framework

In this section, the hierarchical Bayesian framework of Chapter 2 is extended to model the spatio-temporal evolution of the Argo data. Let $H_{\text{obs}}(\mathbf{x})$ denote the observed spatio-temporal field of heat content values. This differs from the "true" unknown heat content field $H(\mathbf{x})$ since the former may include measurement error or small-scale variation. Specifically, let

$$H_{\text{obs}}(\mathbf{x}) \overset{\text{i.i.d.}}{\sim} N(H(\mathbf{x}), \sigma^2(\mathbf{x}))$$

where $\sigma^2(\mathbf{x})$ is the nugget variance. Under this general formulation, $\sigma^2(\mathbf{x})$ can be taken to be varying in both space and time, however the forthcoming construction will restrict this and the other correlation parameters to vary only over space.

3.3.1 Temporal Correlation Model

Let $\theta_{\text{lat}}(\mathbf{x})$, $\theta_{\text{lon}}(\mathbf{x})$, and $\theta_{\text{time}}(\mathbf{x})$ represent spatially-varying latitudinal, longitudinal, and temporal range parameters respectively; these fields will be collectively referred to as $\boldsymbol{\theta} = (\theta_{\text{lat}}, \theta_{\text{lon}}, \theta_{\text{time}})$. At a fixed spatio-temporal location $\mathbf{x}$, the kernel distance between spatio-temporal points $\mathbf{x}$ and $\mathbf{u}$ is defined as

$$d_{\text{st}}(\mathbf{x}, \mathbf{u}; \boldsymbol{\theta}) = \sqrt{\frac{d_{\text{euc}}(x_{\text{lat}}, u_{\text{lat}})^2}{\theta_{\text{lat}}(\mathbf{x})} + \frac{d_{\text{gc}}(x_{\text{lon}}, u_{\text{lon}})^2}{\theta_{\text{lon}}(\mathbf{x})} + \frac{d_{\text{euc}}(x_{\text{time}}, u_{\text{time}})^2}{\theta_{\text{time}}(\mathbf{x})}}.$$

Putting this distance function into the kernel convolution construction of Paciorek and Schervish [2006b] yields the following covariance function between spatio-temporal points $\mathbf{x}$ and $\mathbf{y}$:

$$k_{\text{st}}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}, \phi) = \sqrt{\phi(\mathbf{x})\phi(\mathbf{y})} \int_{\mathbf{u} \in \mathbb{CL} \times \mathbb{R}} \exp\left(-d_{\text{st}}(\mathbf{x}, \mathbf{u}; \boldsymbol{\theta})^2 - d_{\text{st}}(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta})\right)^2 d\mathbf{u}$$

where ϕ is the spatially-varying variance field. Since the squared distance terms d_{st} can be decomposed as the additive sum of distance terms in the three dimensions, the convolutions can be computed in each dimension separately. Specifically, convolutions in the latitudinal and temporal dimensions can be computed exactly using the formulas of Paciorek and Schervish [2006b], and the convolutions in the longitudinal dimension can be computed using the Gaussian approximation method described in Section 2.4.

The underlying heat content field $H(\mathbf{x})$ is modeled as a Gaussian process constructed from the kernel k:

$$H(\mathbf{x}) \sim GP(\mu(\mathbf{x}), k_{\text{st}}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}, \phi)) \tag{3.3}$$

where μ will be defined in Section 3.3.2. Let $\boldsymbol{\rho} = \{\theta_{\text{lat}}, \theta_{\text{lon}}, \theta_{\text{time}}, \phi, \sigma\}$ denote the set of parameter fields excluding the mean-field. For each $\rho \in \boldsymbol{\rho}$, let

$$\log(\rho(\mathbf{x})) \sim GP(\mu_\rho, \phi_\rho \exp(-d_{\text{cyl}}(\mathbf{x}, \mathbf{y})/\theta_\rho)) \tag{3.4}$$

where μ_ρ, ϕ_ρ, and θ_ρ are hyper-parameters which differ for each field. Note that d_{cyl}, defined in Section 2.3.1, is used as opposed to d_{st}; since this distance metric only takes into account the spatial components of $\mathbf{x}$ and $\mathbf{y}$, Equation (3.4) constrains the parameter fields ρ to be constant over time. These parameter fields are taken to vary only spatially primarily to avoid adding additional complexity to the model, particularly given the computational burden of the MCMC procedure with the large number of observations present here. There may be reasons to think that some of the parameters would vary seasonally, and if inference on spatio-temporally varying correlation parameters could be done in an efficient way this could be a potential extension for future work. However it is likely that such an extension would only have a minor effect on inference of the trend field.

3.3.2 Seasonally-Varying Mean-field

Unlike the correlation parameters, there is a stronger reason to think that the mean-field would vary seasonally, and given the stronger relationship between the mean-field and the

trend field it is important for a spatio-temporal model of ocean heat content to take seasonality into account. Furthermore, the patterns of seasonality would not be expected to be constant over the domain; in particular one would expect heat content to peak in the late Northern Hemisphere summer or fall in the higher latitudes and peak in the late Northern Hemisphere winter or early spring in the Southern Hemisphere. To this aim, the spatially-varying mean-field of Chapter 2 is extended to a field varying over both space and time, employing a "circular" correlation model over the temporal dimension in order to produce a field that is smooth at calendar-year endpoints.

To represent the temporal dimension, let $\text{doy}(x_{\text{time}}) \in [0, 366]$ denote the "day-of-year" for a particular time point x_{time}. Let $d_{\text{doy}}(x_{\text{time}}, y_{\text{time}}) = \min(|\text{doy}(x_{\text{time}}) - \text{doy}(y_{\text{time}})|, |365 - \text{doy}(x_{\text{time}}) + \text{doy}(y_{\text{time}})|)$ denote the "circular" distance over day-of-year. Then, analogously to d_{st}, the spatio-day-of-year distance can be defined as

$$d_{\text{s-doy}}(\mathbf{x}, \mathbf{y}; \theta_{\text{lat}}, \theta_{\text{lon}}, \theta_{\text{time}}) = \sqrt{\frac{d_{\text{euc}}(x_{\text{lat}}, y_{\text{lat}})^2}{\theta_{\text{lat}}} + \frac{d_{\text{gc}}(x_{\text{lon}}, y_{\text{lon}})^2}{\theta_{\text{lon}}} + \frac{d_{\text{doy}}(x_{\text{time}}, y_{\text{time}})^2}{\theta_{\text{time}}}}.$$

where unlike in the definition of d_{st} the spatial dependence on correlation length-scale parameters is omitted since this distance function will only be used in the context of a stationary Gaussian process. Specifically, define the spatio-temporal mean-field as

$$\mu_{\text{st}}(\mathbf{x}) \sim \text{GP}(\mu_{\mu_{\text{st}}}, \phi_{\mu_{\text{st}}} \exp(-d_{\text{s-doy}}(\mathbf{x}, \mathbf{y}; \theta_{\mu;\text{space}}, \theta_{\mu;\text{space}}, \theta_{\mu;\text{time}}))$$

where $\theta_{\text{lat};\mu} \equiv \theta_{\text{lon};\mu} \equiv \theta_{\text{space};\mu}$ produces a process that is spatially isotropic with a separate correlation length-scale in the temporal dimension. While isotropic Gaussian processes are used for the parameter fields, the underlying model for the ocean heat content field remains anisotropic. This prior forces μ_{st} to be constant over each calendar year and to vary smoothly over the calendar-year endpoints.

Similar to Chapter 2, the center of the Gaussian process for ocean heat content will be the sum of the year-constant mean-field and a spatially-varying trend field. Here, similar to Chapter 2, the trend-field β is taken to be spatially-varying only:

$$\beta(\mathbf{x}) \sim GP(\mu_\beta, \phi_\beta \exp(-d_{\text{cyl}}(\mathbf{x}, \mathbf{y})/\theta_\beta)).$$

With these two components, the Gaussian process mean of Equation 3.3 is defined as

$$\mu(\mathbf{x}) = \mu_{\text{st}}(\mathbf{x}) + \beta(\mathbf{x})(x_{\text{time}} - x_{\text{origin}})$$

where unlike in Chapter 2 the trend parameters are over units of days rather than years. Here, x_{origin} is defined as January 1, 2007.

The hyper-parameters in Equation 3.4 and above are estimated through obtaining a "first-guess" estimate of the parameter fields through a stationary moving-window approach analogous to the procedure of Section 2.6.1; the results of this procedure are described in Section 3.4. Also similarly to Chapter 2 these fields will be represented and sampled through Gaussian process basis functions. Specifically, for each $\rho \in \boldsymbol{\rho}$, let $\mathbf{b}_\rho \sim N(0, I_{n_{\kappa;\text{space}}})$ denote basis values on a set of spatial knot locations denoted $\kappa_{\text{space}} \in \mathbb{R}^{n_\kappa \times 2}$. Define $\mathbf{b}_\beta$ analogously, and for the mean-field μ_{st}, define $\mathbf{b}_{\mu_{\text{st}}} \sim N(0, I_{n_{\kappa_{\text{st}}}})$ where $\kappa_{\text{st}} = \kappa_{\text{space}} \times \kappa_{\text{time}}$. Here, κ_{st} is taken to consist of 253 knot points evenly spaced over the spatial mask at every 16 degrees in the longitudinal direction and every 8 degrees in the latitudinal direction, and κ_{time} consists of time points distributed bi-monthly. For each $\rho \in \boldsymbol{\rho}$ let the following hold for each location $\mathbf{x}$:

$$\rho(\mathbf{x}) = \log(\mu_\rho + \phi_\rho \Sigma_{\mathbf{x},\kappa_{\text{st}};\theta_\rho} \Sigma_{\kappa_{\text{space}},\kappa_{\text{space}};\theta_\rho} \mathbf{b}_\rho). \tag{3.5}$$

For $\mathbf{b}_\beta$ and $\mathbf{b}_{\mu_{\text{st}}}$, equations analogous to (3.5) hold by removing the logarithmic link function and substituting κ_{st} for μ_{st}.

To evaluate the posterior of the parameter fields, Markov Chain Monte Carlo (MCMC) with Metropolis-Hastings-within-Gibbs steps is used on $\mathbf{b}_\rho$ for each $\rho \in \boldsymbol{\rho}$. The procedure is analogous to that described in Section 2.3.3 of Chapter 2, to which the reader is directed for specific details.

3.4 Initial Configuration of Correlation Parameter Fields

To obtain the initial configuration for the sampler, and to estimate the hyper-priors for the Gaussian process priors assigned to each of the parameter fields, a moving-window approach is used to obtain estimated parameter fields to which Gaussian processes can be fit using maximum likelihood estimation. Specifically, for each grid-point of a $4^\circ \times 4^\circ$ latitude-longitude partition of the domain, parameters of a stationary Gaussian process were estimated from data within a $20^\circ \times 20^\circ$ window. This involves estimating parameters $\theta_{\text{lat}}, \theta_{\text{lon}}, \theta_{\text{time}}, \phi, \sigma, \beta$, and μ_{doy} from the data for each window, where μ_{doy} is an alteration of μ_{st} that is spatially stationary but contains the same structure of seasonal variation. The values of μ_{doy} at the origin, defined as January 1, 2007, is used to obtain the spatial hyper-parameters for μ_{st}. The temporal hyper-parameter for μ_{st} is taken to be fixed at $\theta_{\mu;\text{season}} \equiv 60^2$, which corresponds to a 0.05 effective correlation of 100 days, in order to achieve realistic smoothness in the estimated spatial pattern. This procedure yields a spatially-varying field of stationary parameter estimates for each of the parameters. To estimate the hyper-parameters $\{\mu_\rho, \phi_\rho, \theta_\rho\}_{\rho \in \boldsymbol{\rho}}$, Gaussian processes were fit to each of the respective gridded fields of log-transformed parameter estimates. For the mean and trend fields, spatial range parameters were estimated under the constraint $\theta_{\mu;\text{space}} \equiv \theta_\beta$ in order to ensure that the two fields that compose μ have the same spatial range. Furthermore, μ_β was constrained to be zero in order to avoid biasing the value of the overall trend. The estimated hyper-parameters obtained are displayed in Table 3.1.

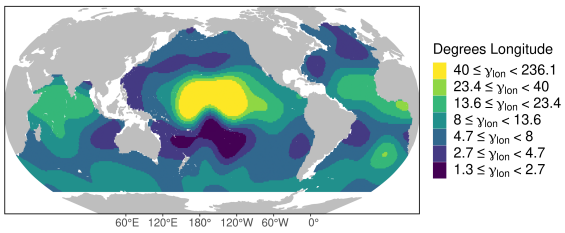
The estimated hyper-parameters displayed in Table 3.1 were used along with standard kriging formulas to obtain the basis values at the spatial knot points κ_{space}. In order to visualize these configurations, these basis values were then kriged again onto a $1^\circ \times 1^\circ$ latitude-longitude grid on the domain, with results displayed in Figure 3.2. Here it can be seen longitudinal correlations are larger than latitudinal correlations in most areas, which is to be expected as the climate processes which influence ocean heat content vary more considerably

	Effective Range	$q_{.25}$	$q_{.5}$	$q_{.75}$	Units
γ_{lat}	27.09°	2.82	3.61	4.61	Degrees
γ_{lon}	29.09°	4.82	7.74	12.44	Degrees
γ_{time}	37.60°	119.48	173.17	250.97	Days
σ^2/ϕ	24.39°	0.04	0.07	0.16	Unitless
$\sqrt{\phi}$	40.16°	3.52	4.89	6.80	GJ/m^2
μ_{st}	44.15°	34.20	45.27	56.34	GJ/m^2
β	44.15°	-0.10	0.00	0.10	(GJ/m^2)/year

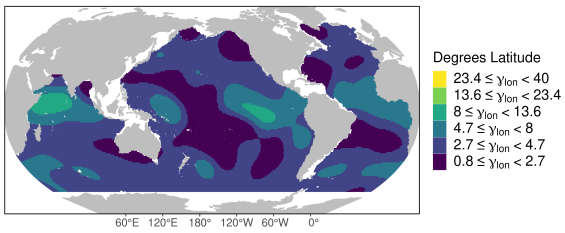
Table 3.1: Hyperparameters obtained from maximum likelihood estimation on fields of stationary parameters estimated from a $20° \times 20°$ moving window procedure. For each field the 5th, 50th, and 95th quantiles of the respective prior distributions are displayed. The range hyperparameter and the correlation length scale fields γ_{lat}, γ_{lon}, and γ_{time} are reported in terms of effective range, defined as the distance at which the correlation is 0.05.

over latitude than over longitude. Longitudinal, and to a lesser degree latitudinal, correlation lengths are noticeably larger around the Equatorial regions, particularly in the Pacific. The process standard deviation field $\sqrt{\phi}$ displays higher values around the major currents, in particular the Gulf Stream of the North Atlantic, the Kuroshio Current off the coast of Japan, the South Atlantic current off of Argentina, and the Arctic circumpolar current in the Southern Ocean. The nugget standard deviation also exhibits higher values in the area of these currents as well as noticeably high values in the South Pacific. The temporal effective correlation lengths are generally on the scale of several months. There is a notable area of temporal correlation lengths of less than a month in the western equatorial Pacific, and there are regions in the Northern Pacific and the Southern Pacific where temporal correlation lengths are more than a year.

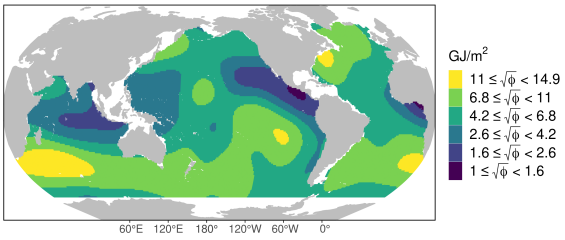
This spatio-temporal initial configuration can be compared to the spatial-only initial configuration displayed in Figure 2.7. The pattern of the process variance having higher values



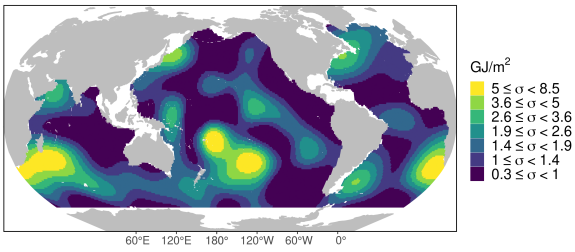
(a) Effective longitudinal range γ_{lon}



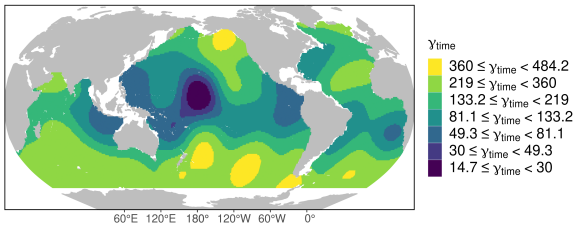
(b) Effective latitudinal range γ_{lat}



(c) Process standard deviation $\sqrt{\phi}$



(d) Nugget standard deviation σ



(e) Effective temporal range γ_{time}

Figure 3.2: Initial parameter fields obtained from fitting Gaussian processes to estimated parameter fields obtained from a moving-window approach.

around major currents is consistent between the spatial-only and spatio-temporal models. This indicates that the variance in these regions persists when accounting for temporal correlations and seasonal patterns in the larger dataset. While both models have higher longitudinal and latitudinal correlation length scales near the equator, the pattern is less robust in the spatio-temporal case, with the central equatorial Pacific having lower correlation scales than in the spatial-only initial configuration. Longitudinal correlations in the equatorial regions are even less prominent in the posterior distribution to be seen in Figure 3.3. This could perhaps be caused by climate phenomena with seasonal features that are not present in the January-only spatial treatment, but are present and modeled in the spatio-temporal context. An example would be ENSO, which causes spatially-coherent anomalies in the equatorial Pacific but is known to peak in the Northern Hemisphere winter [Glantz et al., 2001]. As such, the estimated correlation patterns in the spatial-only treatment may not be present in other months. Furthermore, such seasonal phenomenon would be modeled more precisely with the seasonally and spatially varying mean-field here, which could account for some of the correlation in the anomalies and lead to smaller ranges. Another discrepancy between the spatial-only and spatio-temporal initial conditions can be seen in the process and nugget variance fields, which exhibit high values in the south-eastern and south-western Pacific regions in the spatio-temporal configuration. These regions are unremarkable in the spatial-only configuration. This discrepancy could be caused by the same dynamics as the differences in the correlation length fields, as it is known that the variance and range parameters are closely related in Gaussian process likelihoods [Stein, 1999a]. In other words, it would not be surprising that a change in the properties of the data when adding additional months would manifest in all of the correlation-length and variance fields. Precise identification of the climate feature causing this discrepancy may be of independent interest, and is left here as a suggestion for future research.

3.5 Selection of Vecchia Process Conditioning Set Size

Similar to Section 2.5, a Vecchia process approximation to the full Gaussian process [Guinness and Katzfuss, 2018, Katzfuss et al., 2020c,b, Katzfuss and Guinness, 2021] will be used to render Bayesian inference tractable on the 1,111,023 observations in the dataset. To achieve a sufficiently accurate Vecchia process, the parameter m, which controls the size of the conditioning sets, needs to be re-selected within the new spatio-temporal context. In Chapter 2 a value of $m = 50$ was justified by fitting Gaussian processes with the full Cholesky likelihood to each of eleven regions partitioning the domain and then comparing the resulting integrated values to those obtained from fitting Vecchia processes with varying values of m to the same regions. With the spatio-temporal dataset, however, the full Gaussian process likelihood is too computationally intensive to fit on data restricted to the regions defined in Figure 2.6. Further restricting the regions would lead to areas that would be insufficiently large for estimating the non-stationary parameter fields. Instead, to maintain the full correlation structure of the global model, the study presented here will focus on evaluating the posterior distributions for the mean and trend parameters conditioned on the values of the fully non-stationary correlation structure. This has the downside of assuming a fixed correlation structure rather than re-estimating the parameters, however, given the large amount of time required to fit the fully non-stationary posterior distribution, re-estimating the parameter fields for different values of m would not be practical.

Define the posterior mean and standard deviation of the integrated mean-field conditioned on the set of correlation parameter fields $\boldsymbol{\rho}$ as

$$\left[\int_{x_{\text{lat}},x_{\text{lon}}} \mu_{\text{st}}(x_{\text{lat}}, x_{\text{lon}}, x_{\text{origin}})\, dx_{\text{lat}}\, dx_{\text{lon}} \middle| H_{\text{obs}}, \boldsymbol{\rho}\right] \equiv N(\mu_{\text{post}}, \sigma_{\mu;\text{post}})$$

where x_{origin} denotes January 1, 2007. For the trend field, analogously define

$$\left[\int_{x_{\text{lat}},x_{\text{lon}}} \beta(x_{\text{lat}}, x_{\text{lon}})\, dx_{\text{lat}}\, dx_{\text{lon}} \middle| H_{\text{obs}}, \boldsymbol{\rho}\right] \equiv N(\beta_{\text{post}}, \sigma_{\beta;\text{post}})$$

These quantities can be calculated directly using the formulas of Equation (2.5) in Section

	μ_{post}	$\sigma_{\mu;\text{post}}$	β_{post}	$\sigma_{\beta;\text{post}}$
5	3.12×10^{-4}	0.049	8.30×10^{-3}	0.046
10	1.84×10^{-4}	0.027	4.53×10^{-3}	0.025
20	1.33×10^{-4}	0.014	6.95×10^{-3}	0.013
30	8.67×10^{-5}	9.76×10^{-3}	5.34×10^{-3}	9.41×10^{-3}
40	7.11×10^{-5}	7.46×10^{-3}	4.36×10^{-3}	7.31×10^{-3}
50	6.85×10^{-5}	5.93×10^{-3}	5.62×10^{-3}	5.89×10^{-3}
60	4.61×10^{-5}	4.95×10^{-3}	3.66×10^{-3}	4.87×10^{-3}
70	3.58×10^{-5}	4.15×10^{-3}	3.10×10^{-3}	4.10×10^{-3}
80	3.05×10^{-5}	3.62×10^{-3}	2.92×10^{-3}	3.55×10^{-3}
90	2.68×10^{-5}	3.20×10^{-3}	3.14×10^{-3}	3.09×10^{-3}
100	2.51×10^{-5}	2.83×10^{-3}	2.73×10^{-3}	2.77×10^{-3}
200	2.40×10^{-6}	1.17×10^{-3}	6.60×10^{-4}	1.12×10^{-3}
Reference: 500	0	0	0	0

Table 3.2: Fractional errors of posterior means and standard deviations for the mean and trend fields evaluated using Vecchia processes with increasing values of m. Errors are calculated against reference level of $m = 500$.

2.3.3. In order to select m before beginning the MCMC sampling procedure, $\boldsymbol{\rho}$ is taken to be the initial condition parameter fields obtained in Section 3.4.

The posterior distribution parameters defined above will be denoted with super-script (m) when calculated using the corresponding Vecchia process. Values of $\mu^{(m)}_{\text{post}}$, $\sigma^{(m)}_{\mu;\text{post}}$, $\beta^{(m)}_{\text{post}}$ and $\sigma^{(m)}_{\beta;\text{post}}$ were computed for $m \in \{5, 10, 20, \ldots, 100, 200\}$ and $m = 500$. Since it is not computationally possible to evaluate the full Cholesky decomposition with over one million observations, $m = 500$ was the highest value of m evaluated; with this value, it took over eight hours with 30 cores on a cluster with 512GB of memory to evaluate the posterior distributions. For each value of m, fractional errors for the mean and slope fields were

calculated using $m = 500$ as the reference level, yielding $\frac{|\mu_{\text{post}}^{(m)} - \mu_{\text{post}}^{(500)}|}{|\mu_{\text{post}}^{(500)}|}$, $\frac{|\sigma_{\mu;\text{post}}^{(m)} - \sigma_{\mu;\text{post}}^{(500)}|}{|\sigma_{\mu;\text{post}}^{(500)}|}$, $\frac{|\beta_{\text{post}}^{(m)} - \beta_{\text{post}}^{(m)}|}{|\beta_{\text{post}}^{(500)}|}$, and $\frac{|\sigma_{\beta;\text{post}}^{(m)} - \sigma_{\beta;\text{post}}^{(500)}|}{|\sigma_{\beta;\text{post}}^{(500)}|}$.

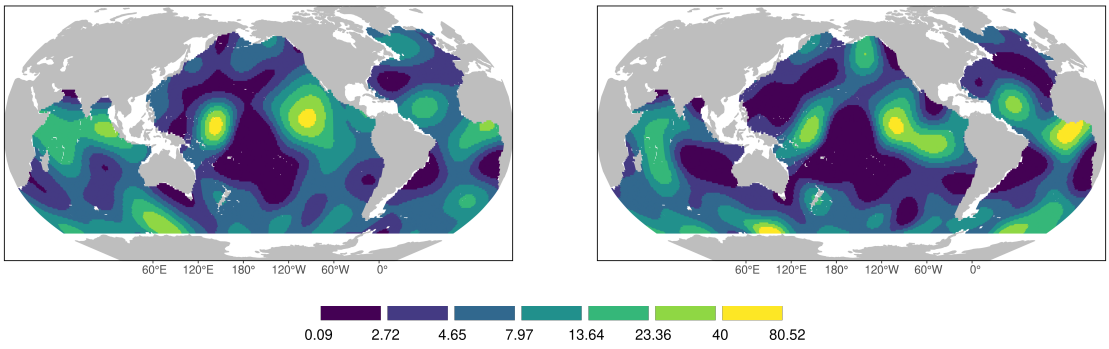
The results of this study are displayed in Table 3.2. It can be seen that the errors for the posterior standard deviation of both the mean and trend fields are higher than the errors for the posterior means. This is unsurprising as the values of the mean and trend fields are fairly constrained by the amount of data available, whereas the uncertainty values are more challenging to estimate accurately. For all four of the evaluated quantities, errors generally descend as the values of m increase. However, after about $m = 50$ the fractional errors are all less than 7.5×10^{-3}, with higher values of m seeing decreasing marginal improvements. As such, $m = 50$ is used for the MCMC results presented in the following section.

3.6 Posterior Distribution Results

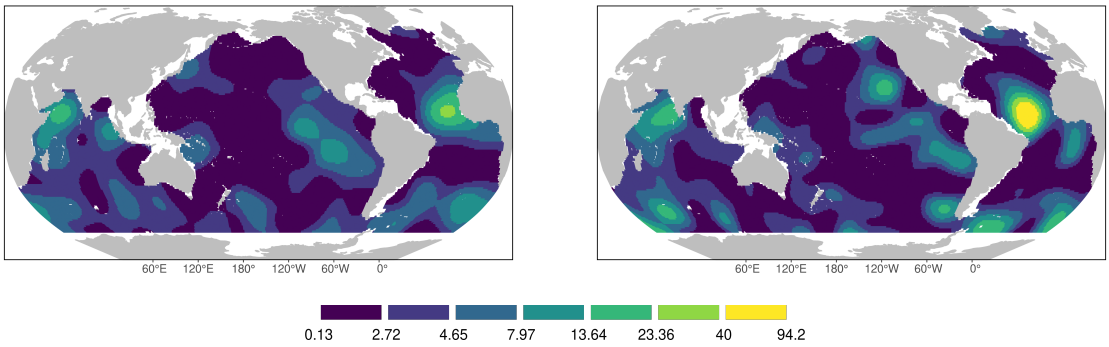
To obtain the posterior distribution of the correlation parameters using MCMC, the sampler was run for increasing lengths of 1,000 iterations until convergence was achieved as indicated by the Heidelburger-Welch test applied to posterior densities of the parameter fields [Heidelberger and Welch, 1981]. This procedure required 3,000 iterations, of which the first 2,500 iterations were removed as a burn-in period. These 3,000 iterations took 528 hours (around 22 days) running on 30 cores of a compute server with 528GB of memory. This section will proceed to present the results from the posterior distribution, including the covariance parameter fields, the spatially and seasonally varying mean-field, the ocean heat content values by month, and the spatially-varying trend field.

3.6.1 Posterior Parameter Fields

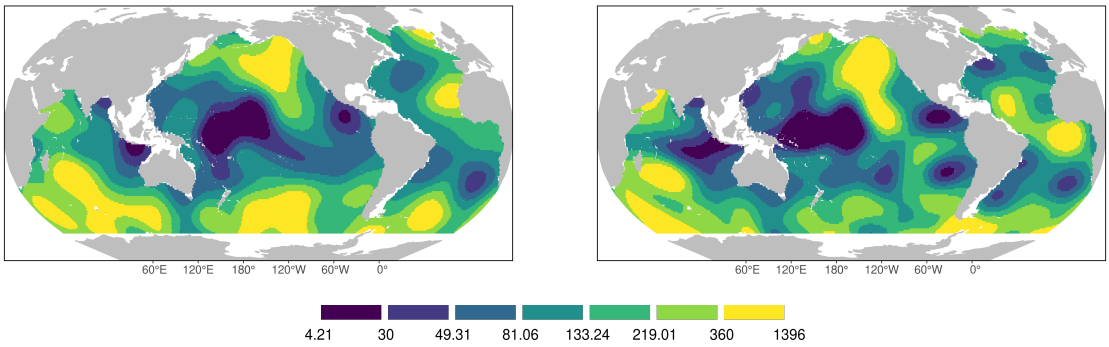
To visualize the posterior distribution, configurations corresponding to the first and ninety-ninth quantiles of the average parameter values were identified for each parameter field. The



(a) Effective longitudinal range γ_{lon} (degrees longitude).

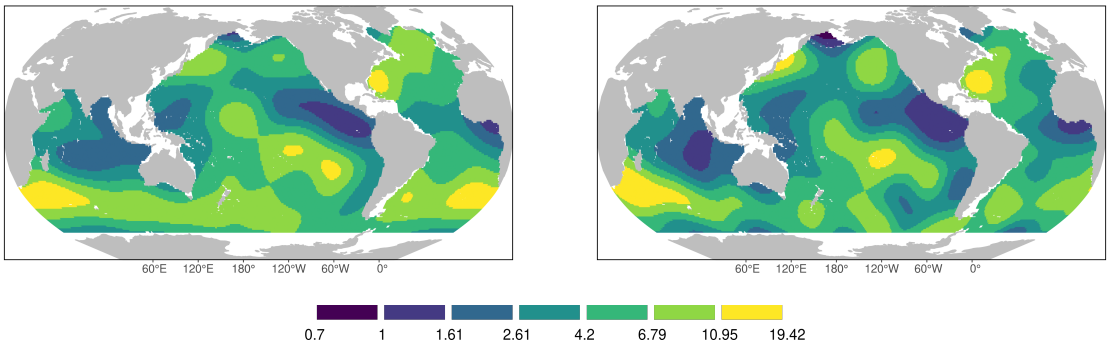


(b) Effective latitudinal range γ_{lat} (degrees latitude).

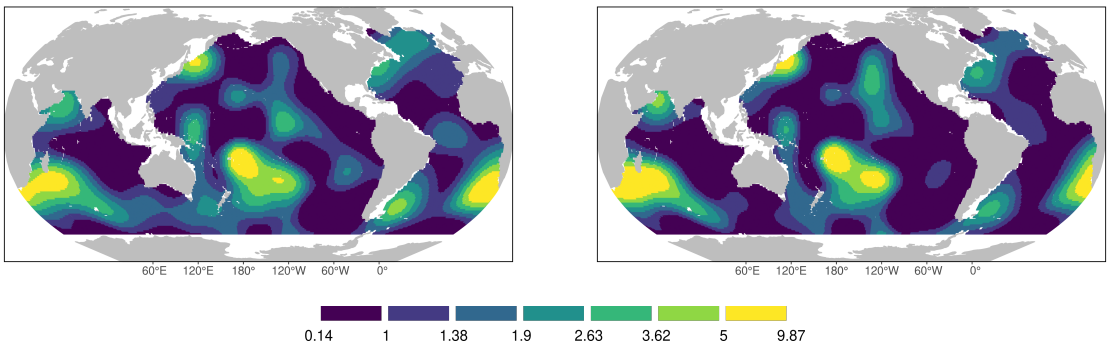


(c) Effective temporal range γ_{time} (days).

Figure 3.3: Samples corresponding to the first (left) and ninety-ninth (right) posterior quantiles for the effective longitudinal, latitudinal, and temporal ranges.



(a) Process standard deviation $\sqrt{\phi}$ (GJ/m^2).



(b) Nugget standard deviation σ (GJ/m^2).

Figure 3.4: Samples corresponding to the first (left) and ninety-ninth (right) posterior quantiles for the process standard deviation and nugget standard deviation.

resulting interpolated fields for longitudinal, latitudinal, and temporal ranges are displayed in Figure 3.3. For longitudinal ranges, it can be seen that the behavior of the high-range region in the equatorial Pacific has decreased in the center and moved both east and west in comparison to the initial configuration of Figure 3.2a. Variability in the posterior distribution for the longitudinal range field can be seen in this equatorial Pacific region, as well as in the equatorial Atlantic and in the Southern Ocean. For latitudinal ranges, Figure 3.3b exhibits higher values in the equatorial Atlantic when compared to the initial configuration, as well as noticeable variability. Variability in latitudinal ranges can also be observed in the eastern equatorial Pacific and the Southern Ocean. For temporal ranges, values in the posterior appear to be generally larger than those in the initial condition, with variability particularly high in the Southern Ocean.

The posterior distributions of process variance and nugget variance are displayed in Figure 3.4. For the process variance, Figure 3.4a shows general patterns similar to the initial configuration. Larger posterior variability is particularly notable around the location of the highly variable Antarctic Circumpolar current south-west of Africa. Overall, the process variance and nugget variance displayed in Figure 3.4 appear to exhibit smaller levels of posterior variability than the other parameters. This is indicative of the fact that the variance parameters are fairly constrained by the amount of the data present, although larger posterior uncertainty is in general larger in the highly variable current regions. It should be noted that a common feature between all of the posterior distributions is the higher variability in the Southern ocean. This is not surprising when considering the spatial distribution of floats in Figure 3.1, where despite the large increase in float coverage over time, the data coverage rates in the Southern ocean remain low in later years.

3.6.2 Posterior Mean, Trend, and OHC Values

After obtaining the posterior distributions for the correlation parameter fields using MCMC, the posterior distribution of the spatio-temporal mean and trend fields can be evaluated

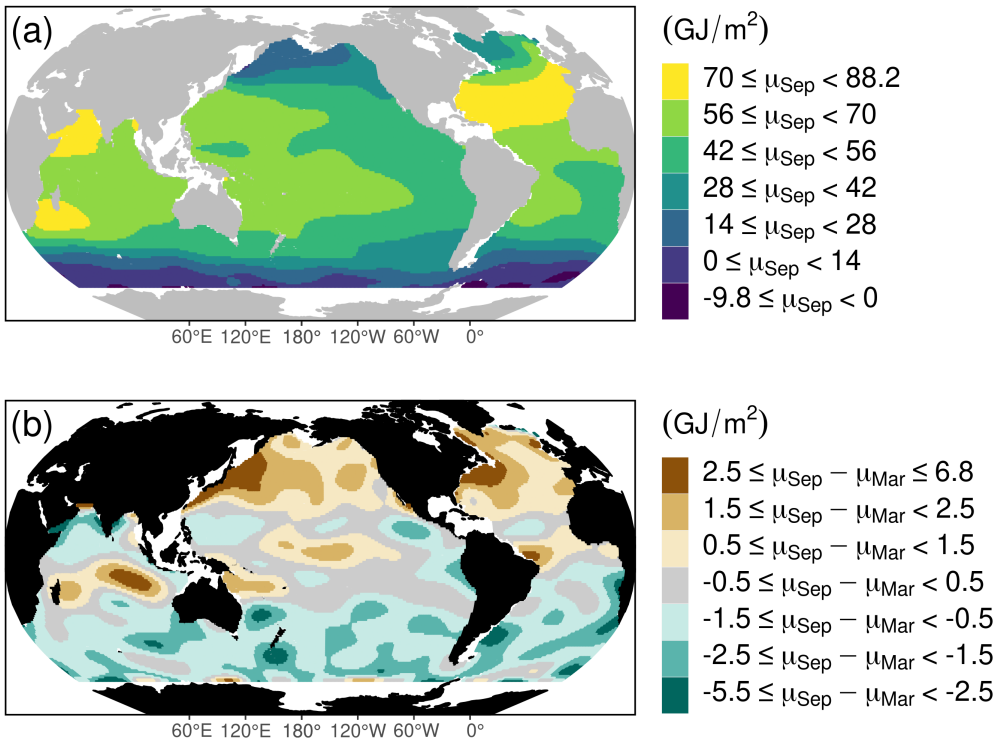


Figure 3.5: Mean-fields for September (a) and the difference between September and March (b) showing spatially-varying seasonality.

directly using Equation (2.5) in Section 2.3.3. Figure 3.5 shows the mean-field for the MAP configuration evaluated at September 15. To get a sense of the spatial variation of seasonality, Figure 3.5(b) displays the difference between the September 15 mean-field minus the mean-field evaluated at March 15. As the ocean absorbs heat slowly throughout local summer, the warmest ocean heat content values would be expected to occur around the month of September in the Northern Hemisphere. Conversely, March would be expected to see the warmest ocean heat content values in the Southern Hemisphere. This intuition is confirmed in Figure 3.5(b), where it can be seen that September is generally warmer than March in the Northern Hemisphere and is generally colder than March in the Southern Hemisphere. Seasonal differences are most pronounced around currents, and in particular the Kuroshio current east of Japan and the North Atlantic current east of the United States.

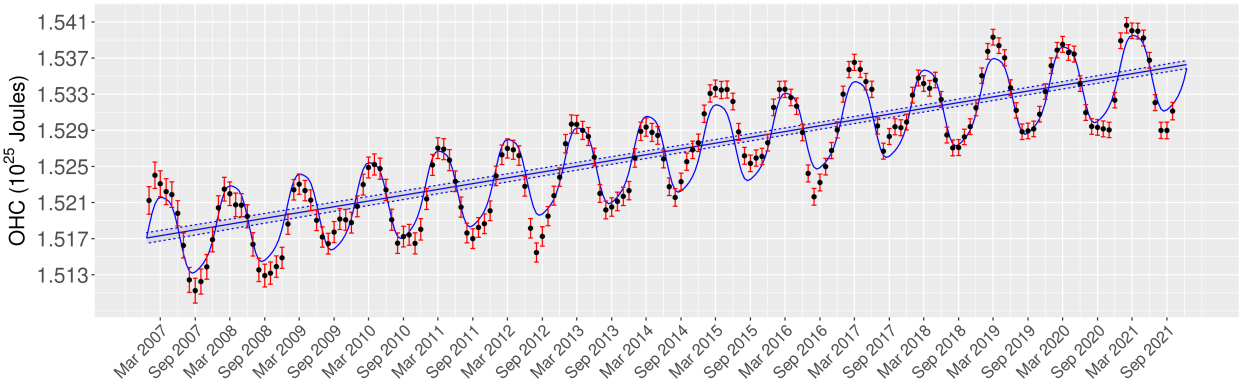


Figure 3.6: OHC by month with credible intervals. The 95% credible interval for the trend is $(12.59, 13.39) \times 10^{21}$J/year.

Figure 3.6 displays the posterior distribution for globally-integrated ocean heat content evaluated at each month of the time range with 95% credible intervals. The distinction between confidence intervals and credible intervals made in Figure 2.9 are ignored here, as they are often difficult to distinguish in the full spatio-temporal setting. It can be seen that the width of the credible intervals for the ocean heat content values decrease over time as the number of floats increases. Also displayed is the posterior mean of the globally-integrated mean-field. The seasonal structure of ocean heat content is clearly visible in this figure, with the warmest values occurring in late March. This is due to the fact that March sees the warmest ocean heat content values in the Southern Hemisphere, which contains a larger percentage of the global ocean than the Northern Hemisphere.

Figure 2.9 additionally displays the mean globally-integrated trend over time and its corresponding 95% credible interval. It can be seen here that the massive increase in data produces a notably narrower credible interval for the trend than in Figure 2.9. In particular, here the 95% credible interval for the globally integrated trend is (12.59,13.39) ZJ/year, whereas in the results of Chapter 2 the interval was (4.66,16.03) ZJ/year. This corresponds to reduction in the posterior standard error by a factor of 15. This is larger than would be expected from the increase in the amount of data alone, as if the $1/\sqrt{n}$ convergence rate of

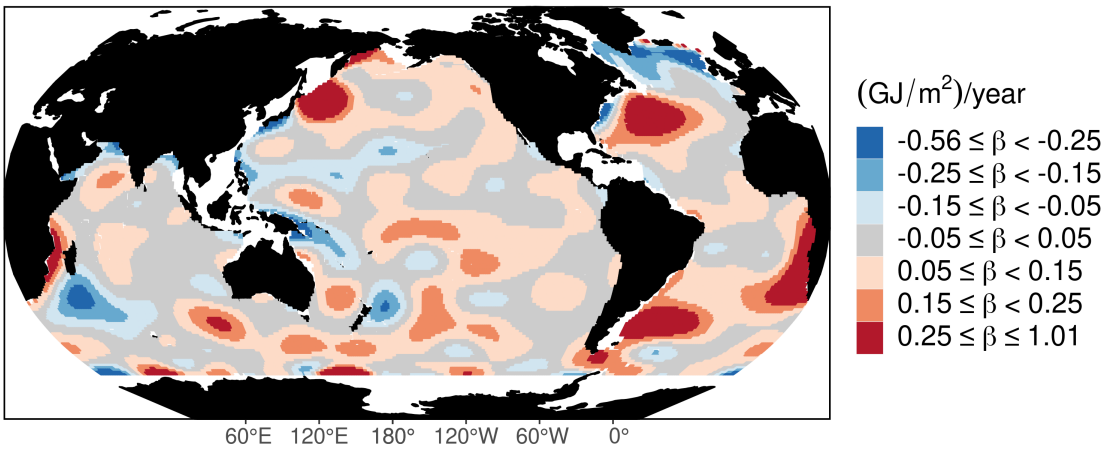


Figure 3.7: Posterior mean for the trend field β in units of (GJ/m^2)/year.

the variance of a standard normal distribution is assumed, increasing the number of data points from 47,513 to 1,147,710 would suggest a reduction in the standard error by a factor of 5. This indicates that adding the spatial components to the model is adding additional confidence to the multi-year trend than would be expected from the amount of data alone. This can be attributed to three sources. The first and possibly the largest effect is that, with the addition of the temporal correlation structure, values at locations that may not have any nearby observations in a particular month now can use information from anomalies in nearby months. This will reduce the interpolation uncertainty over the single-month version whenever the effective temporal range is greater than a month, which as seen in Figure 3.3c is true for most of the ocean heat content domain. The second effect is that the addition of floats has increased the spatial coverage of the observations as can be seen in Figure 3.1. This increase in coverage will lead to a reduction of interpolation uncertainties in more recent years. The third effect is the inclusion of seasonal variation in the mean-field, which will reduce the values of the anomalies by modeling the albeit small amount of within-month seasonal variation.

The posterior mean of the spatially-varying trend field is displayed in Figure 3.7. Since

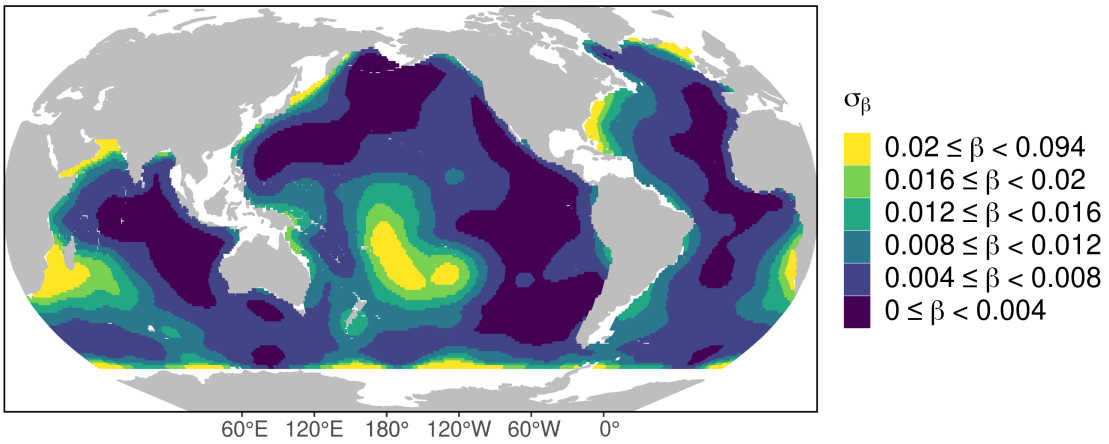


Figure 3.8: Posterior standard deviations of the trend field β in units of (GJ/m^2)/year.

the posterior means are more than three standard deviations away from zero in 84% of the grid-cells, the posterior probability of a positive trend visualization from Figure 2.10 is omitted in lieu of the map of posterior standard deviations in Figure 3.8. From Figure 3.8 it can be seen that the areas of the trend field with the highest uncertainty generally correspond to the high-variance regions displayed in Figure 3.4.

When comparing these results to Figure 2.10, some areas such as the Atlantic Ocean east of North America and south-east of South America have warming trends in both the January-only fit and in the all-years fit. The clear warming and cooling trends in the equatorial Pacific in the January-only fit are not as prominent in the all-years fit. This likely is due to the fact that ENSO's somewhat seasonal features has given it an outsized influence on the equatorial Pacific region in the January-only results from Chapter 2.

3.7 Conclusion

The results in this chapter has shown that extending the January-only model from Chapter 2 to all months and to additional years allows for ocean heat content to be estimated with

a much higher degree of accuracy. In particular, the incorporation of temporal structure in the covariance structure of the anomalies and incorporating seasonality in the mean-field generates results with higher confidence than would be expected just from the increase in the amount of data alone. Another notable change when extending the model to the spatio-temporal domain is the change in the posterior distributions for the latitudinal and longitudinal ranges as well as in the marginal variance and nugget variance parameters in the central Pacific. This suggests that allowing the correlation parameters to vary seasonally could present a meaningful improvement to the statistical modeled presented here. Since the estimation of the posterior distribution of ocean heat content is fairly constrained by the amount of information present in the data, further improvements in modeling the covariance structure of the ocean heat content field will likely have only a small effect on the strength of the ultimate conclusions. It should also be noted that the trend found here is broadly consistent with other entries in the literature [Cheng et al., 2019, Resplandy et al., 2019], although a more in-depth analysis of how the mean estimates and uncertainty levels compare between different approaches, and how these estimations influence estimates of Equilibrium climate sensitivity, would be an interesting extension in future work.

CHAPTER 4

Representing Uncertainty in Estimating the the Covariance Matrix in Optimal Fingerprinting

The detection of the influence of increasing greenhouse gas concentrations on observed temperatures requires the estimation of the forced signal as well as the internal variability covariance matrix in order to distinguish between their contributions. While previous approaches have taken into account the uncertainty linked to estimation of the forced signal, there has been less focus on uncertainty in the internal covariance matrix describing natural variability, despite the fact that the specification of this covariance matrix is known to meaningfully impact the results. In this chapter, a covariance matrix parameterization using Laplacian basis functions is proposed. This parameterization, unlike traditional methods which are based on principal component basis representations, does not require the basis vectors themselves to be estimated from climate model data. The Laplacian parameterization is incorporated into a Bayesian regression approach which allows for the propagation of uncertainty in estimating the covariance structure to the final results. To select the number of basis functions, a separate Bayesian model using a χ^2 re-parameterization of the likelihood is proposed with the aim of more accurately representing the statistical properties of the covariance matrix. The final proposed statistical approach involves fitting the two Bayesian models simultaneously in an iterative fashion. A validation study using CMIP6 climate model ensembles shows that the proposed approach achieves better-calibrated coverage rates of the true regression parameter than principal component-based approaches. When applied to HadCRUT observational data the proposed approach produces higher levels of confidence in detecting anthropogenic warming, with lower variability over the choice of climate models, than a principal-component based approach.

4.1 Introduction

While it is well-established that increasing greenhouse gas concentrations will lead to warming surface temperatures [Arrhenius, 1896, Masson-Delmotte et al., 2021], it is difficult to quantify the extent to which changes throughout the observational record can be attributed

to human influence. This issue is addressed through climate change detection and attribution (D&A) research, where a combination of General Circulation Model (GCM) output and statistical techniques are used to separate the anthropogenic signal from natural variability in the observational record. "Optimal fingerprinting" is in particular concerned with identifying the relationship between global "forcing" patterns. Forcing is defined as any change which influences the radiative balance of the atmosphere, and while the primary forcing under consideration here is increasing greenhouse gas concentrations, other forcings include changing aerosol concentrations and changes in solar radiation due to natural cycles. Optimal fingerprinting frameworks generally proceed by regressing the observed pattern of a particular climate variable onto a set of forcings estimated using climate models, with the residuals attributed to natural variability. Statistical inference on regression coefficients can then be used to assess the extent to which the forced signal can be "detected" in the observations and the degree to which observed trends can be "attributed" to the forcing. The regression framework has formed the conceptual foundation of optimal fingerprinting since it was introduced by Hegerl et al. [1996], who used generalized least squares (GLS) to give the residuals a covariance structure that can be informed by GCM simulations of natural variability.

The reliability of optimal fingerprinting conclusions depends crucially on their ability to accurately represent uncertainty in the regression coefficient. This is complicated by the fact that various sources of uncertainty are present in the optimal fingerprinting procedure, including observational uncertainty, uncertainty in estimating the forced component, and uncertainty in representing the natural variability covariance matrix. In light of this, while the GLS framework takes into account the uncertainty induced by natural variability itself, there have been efforts in the optimal fingerprinting literature to expand the framework and incorporate additional sources of uncertainty. Allen and Stott [2003] proposed the use of a total least squares inference approach to take into account the uncertainty induced by estimating the forced signal from a limited number of climate model runs. These approaches

assume that the covariance structure is the same under forced and unforced climate scenarios. This assumption is relaxed in the methodology of Huntingford et al. [2006] and Hannart et al. [2014] who allow for separate covariance matrices for different forcing patterns.

A particular challenge with the regression approach is choosing how to represent the covariance matrix. The standard approach is to use a limited number of principal components estimated from the empirical covariance matrix of a set of climate model runs. The reason for truncating the principal components is that, while the first estimated principal components will generally correspond to meaningful patterns of variability, higher-number principal components may not correspond to meaningful patterns and their variances are likely to be under-estimated. The reason for this is that lower-variability directions of the sub-spaces are less likely to be represented within a finite number of control runs than higher-variability components, which leads to under-estimation of the lower-variance eigenvalues [Allen and Tett, 1999]. Under standard GLS regression formulas, terms are weighted according to the inverse of their variance, which is often called the "precision". Under-estimated variances, or over-estimated precisions, will artificially increase the weight of non-meaningful patterns in the inference. A commonly used way to choose the number of principal components is the residual consistency test of Allen and Tett [1999], which rejects components with variances that are unrealistically small as gauged by residuals from the regression model. Another approach to minimizing the effect of under-estimated variances is through regularization techniques. In particular, Ribes et al. [2009] and Ribes et al. [2012] use the Ledoit and Wolf estimator [Ledoit and Wolf, 2004]. This method can be viewed as obtaining a full-rank covariance matrix from the empirical covariance matrix by scaling upward the variances of higher-number components while "shrinking" the variance of lower-number components under the constraint that the total variance remains the same.

The approaches discussed above have used point estimators of the covariance structure that are estimated prior to model fitting and treated as fixed from the perspective of inference. This precludes the uncertainty from estimating the covariance structure from being

represented in the uncertainty values for the final result. Recent developments have focused on developing detection and attribution methodology that performs covariance estimation and regression in an integrated fashion. Hannart [2016] propose a frequentist integrated inference procedure using Ledoit-Wolf shrinkage of the covariance matrix. Ribes et al. [2017] develop an error-in-variables approach that is not based on the regression framework and is motivated by the incorporation of model uncertainty, meaning the uncertainty induced by the differences between climate models from each other and from reality, as opposed to just estimation uncertainty, which is a product of the limited number of climate model runs.

The two integrated frameworks discussed above take a frequentist view of the optimal fingerprinting problem, where point estimators are developed based on observed data and confidence intervals are constructed based on their sampling distributions. Frequentist frameworks assume that the underlying quantity is fixed but unknown. However, the underlying quantities of the climate system could be more naturally thought of as random variables due to the underlying probabilistic nature of the climate system [Slingo and Palmer, 2011]. A Bayesian philosophy would be a natural way to integrate the various sources of information and uncertainty present into a statistical framework. This is the logic motivating Katzfuss et al. [2017], or KHS17 from here on out, who develop a hierarchical Bayesian approach for optimal fingerprinting. This framework uses the truncated principal component parameterization of the covariance matrix where the component variances and the number of components are treated as variables. In the implementation, the component variances are estimated along with the regression parameters conditioned on each truncation number using MCMC. Then Bayesian model averaging (BMA; Hoeting et al. [1999]) is used to marginalize over the number of components. This allows for the uncertainty in estimating these weights and the number of components in the covariance matrix to be propagated to the final inference results. A limitation to this method is that the principal components are computed in advance and treated as fixed in the inference, so only the uncertainty in estimating the variance coefficients is propagated to the final result.

The goal of integrated D&A approaches is to correctly propagate the various sources of uncertainty to the final result. Better modeling of the uncertainty can lead to higher standard errors than approaches that omit sources of uncertainty or under-estimate component variances. The reliability of these standard errors can be gauged through coverage rates, defined as the percentage of the time that an uncertainty interval contains the true value when the true value is known. Ideally, a 90% confidence interval will contain the true value 90% of the time; if under-coverage is observed it indicates that the model is under-estimating variance or uncertainty. Fortunately, climate model ensembles provide a wealth of data for evaluating coverage rates of statistical methods under the assumption that the observations are truly generated from the appropriate distribution.

Optimal fingerprinting methods have been observed to exhibit under-coverage under various settings. DelSole et al. [2019] show that confidence intervals are biased in a total least squares setting similar to those developed by [Allen and Stott, 2003, Hannart et al., 2014], and propose a bootstrapping method to correct for this bias. Li et al. [2021] demonstrate under-coverage when using the regularized method of [Ribes et al., 2012], and also propose to correct the standard errors using a bootstrapping approach. While these methods are able to achieve coverage rates closer to the nominal values, bootstrapping approaches produce uncertainty values that are no longer valid from the perspective of a probability model on the data. Under-coverage in the uncertainty intervals indicates that there are sources of variability that are not being taken into account, and an ideal approach would identify and statistically model these factors directly. Furthermore, these two papers are concerned with the frequentist properties of the detection and attribution regression coefficients, and such differ from the Bayesian approach considered here.

The chapter aims to develop optimal fingerprinting methodology within a Bayesian framework with improved coverage properties. The proposed approach builds off of the Bayesian framework of KHS17, with two key innovations. The first is the parameterization of the covariance matrix using Laplacian basis functions, rather than principal components, to avoid

the uncertainty induced by estimating principal components from climate model data and to better propagate the uncertainty in estimating the covariance matrix to the final result. In addition, the model-driven component selection approaches using the normal regression likelihood model, such as the BMA approach of KHS17, are driven towards minimizing the residuals rather than accurately modeling the covariance matrix. This produces fits that achieve lower accuracy and misleading coverage when evaluated using climate models. To this end, the second proposed innovation is a separate Bayesian model for the number of components which uses a χ^2 reparameterization of the likelihood accurate representation of the variance terms. The overall proposed approach is to fit these two Bayesian models for the regression and the number of components simultaneously with an iterative process. It will be shown that this approach achieves higher accuracy and better-calibrated coverage rates than when using the principal component parameterization or when using the normal distributional assumption for modeling the number of components.

4.2 Methods

Let $\mathbf{y}$ denote a vector of true climate observations. In general, this vector can be spatial or spatio-temporal, however here it will be assumed that this is a spatial vector giving values of a climate variable at each of n_{grid} grid-points on the sphere. In the spatial-only case, changes in the climate will be represented by taking each point of $\mathbf{y}$ to be the linear regression coefficient of the climate variable of interest over time. The validation study in Section 4.3 and the application in Section 4.4 will be using trends in near-surface air temperatures, specifically the CMIP-standard variable "tas", over the 25-year period 1990-2015. However, optimal fingerprinting methodology can be applied to other climate variables such as precipitation [Lambert et al., 2003, Zhang et al., 2007, Min et al., 2011, Wan et al., 2015], zonally-averaged tropospheric temperatures [Santer et al., 2003, 2013, 2018], precipitable water [Ma et al., 2017, Zhang et al., 2019], oceanic oxygen levels [Andrews et al., 2013], streamflow timing

[Hidalgo et al., 2009], vegetation changes [Mao et al., 2016], temperature extremes [Christidis et al., 2013, Kim et al., 2016, Lu et al., 2016, Seong et al., 2021], and various other climate indices [Timmermann, 1999, Roberts and Palmer, 2012, Hobbs et al., 2015, Weller et al., 2016].

In practice, the true vector $\mathbf{y}$ will not be exactly known. Let $\{\mathbf{y}_1, \ldots, \mathbf{y}_{n_{\text{obs-ens}}}\}$ denote $n_{\text{obs-ens}}$ vectors from an ensemble designed to represent observational uncertainty, for example from the $n_{\text{obs-ens}} = 200$ HadCRUT fields of near-surface air temperatures [Morice et al., 2021]. For each observational ensemble member $\mathbf{y}_k$, let $\mathbf{y}_k \sim \mathcal{N}(\mathbf{y}, W)$ where the covariance matrix W represents observational uncertainty as implied by the ensemble. In the following presentation and results the observational uncertainty matrix W will be ignored in order to focus on the modeling of the covariance structure of natural variability. For the application in Section 4.4, the observational ensemble mean $\frac{1}{n_{\text{obsens}}} \sum_{k=1}^{n_{\text{obsens}}} \mathbf{y}_k$ will be used as a point-estimate of $\mathbf{y}$. However, this is not a major limitation of the approach, as if desired the Bayesian methods discussed can be augmented to include uncertainties obtained from the observational ensemble using the computationally efficient formulas of KHS17.

Let $\mathbf{x}$ denote the true field of climate trends under a "forcing scenario". This forcing scenario here will exclusively refer to historical forcings, which include increasing levels of greenhouse gas concentrations but also aerosals and changes in solar radiation. However, this methodology would still hold if a different forcing scenario were to be used. It should be noted that most treatments of optimal fingerprinting allow for multiple forcing patterns to be included simultaneously, allowing for example greenhouse gas forcing and "natural forcings" to be treated separately. As natural forcings are relatively minor when trends are taken over 25-year periods, this chapter will treat the historical forcing patterns as representative of the effect of greenhouse gases, although extending the approach to include multiple distinct forcings would be straightforward.

Let $\mathcal{P}_c = \{\mathbf{z}_1, \ldots, \mathbf{z}_{n_{\mathcal{P}_c}}\}$ and $\mathcal{H}_f = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n_{\mathcal{H}_f}}\}$ denote sets of $\mathbb{R}^{n_{\text{grid}}}$ vectors of trends under "control" and "forced" simulations respectively, where $1 \leq c \leq N_{\mathcal{P}}$ indexes over the

$N_{\mathcal{P}}$ climate models used under pre-industrial forcing and $1 \leq f \leq N_{\mathcal{H}}$ indexes over the $N_{\mathcal{H}}$ climate models used under historical forcings. The lower-case values $n_{\mathcal{P}_c}$ and $n_{\mathcal{H}_f}$ refer to the number of fields obtained from the c-th control and f-th sets of pre-industrial and historical vectors respectively. Intuitively, $\mathcal{P}$ stands for "pre-industrial", $\mathcal{H}$ stands for "historical", c stands for "control", and f stands for "forced". The dependence of $\mathbf{x}$ and $\mathbf{z}$ on $\mathcal{P}_c$ and $\mathcal{H}_f$ will be omitted when clear from context. It will be assumed that natural variability in the climate system has a multivariate normal distribution, or $\mathbf{z}_i \sim N(0, C_{\mathcal{P}_c})$ for each $1 \leq i \leq n_{\mathcal{P}_c}$ with covariance matrix $C_{\mathcal{P}_c} \in \mathbb{R}^{n_{\text{grid}}, n_{\text{grid}}}$. Let $\mathbf{x}_j \sim N(\mathbf{x}_{\mathcal{H}_f}, C_{\mathcal{P}_c})$ for each $1 \leq j \leq n_{\mathcal{H}_f}$ where $\mathbf{x}_{\mathcal{H}_f}$ is the "true" forced vector for the f-th historical climate model. Note that the natural variability covariance matrix $C_{\mathcal{P}_c}$ is assumed to be unchanged by forcing; this assumption is often made in the literature, although approaches such as Hannart et al. [2014] relax this assumption. The dependence on the climate model used to generate $\mathcal{P}_c$ for covariance matrix C will be dropped when clear from context. The climate models that will be used for generating $\mathcal{P}_c$ and $\mathcal{H}_f$ are displayed in Table 4.1. These will be described in further detail in Section 4.3.1, but it will be noted here that the simulation data was restricted to cases where $n_{\mathcal{P}_c} > 5$ and $n_{\mathcal{H}_f} > 5$ to ensure that there are sufficient numbers of control and historical fields to be used in the procedure.

With this set-up, the key formula driving optimal fingerprinting methodology is

$$\mathbf{y} \sim \beta \mathbf{x} + \epsilon \text{ where } \epsilon \sim N(0, C). \tag{4.1}$$

Here the parameter β is interpreted as the "detection and attribution" parameter. Detection is generally taken to be a rejection of the hypothesis test $H_0 : \beta > 0$, and attribution is taken to be, conditional on detection, whether or not β is statistically indistinguishable from one. In the latter case, $\beta > 0$ indicates that the forcing pattern $\mathbf{x}$ is meaningfully present in the observed response. When β is statistically indistinguishable from one, it indicates that the discrepancy between the observations and the forced pattern is within the range of natural variability and that there is no evidence that other forcings are needed to explain observed trends.

	Number piControl	Number historical
ACCESS-ESM1-5	36	40
AWI-CM-1-1-MR	13	
CanESM5	57	65
CESM2	43	
CESM2-FV2	18	
CESM2-WACCM	17	
CESM2-WACCM-FV2	19	
FGOALS-g3	27	
FIO-ESM-2-0	20	
GISS-E2-1-G	66	46
MIROC6	28	
NESM3	16	
NorCPM1	48	30
NorESM2-MM	15	
SAM0-UNICON	27	
UKESM1-0-LL	61	
GISS-E2-1-H		25
MIROC-ES2L		31
MPI-ESM1-2-LR		29

Table 4.1: Number of pre-industrial control and historical ensemble members used from CMIP6 climate models. Missing numbers indicate that either the scenario was not available for the given climate model or the number of useable 25-year trend fields was not greater than 5.

If $\mathbf{x}$ and C were known, (4.1) could be fit using (GLS), which is a well-understood methodology with closed-form frequentist estimators [Aitken, 1936] and was used by the original approach of Hegerl et al. [1996]. Crucially, however, the fact that these values are not known adds uncertainty to the inference which needs to be taken into account. Nevertheless, the following section will proceed with an overview of the frequentist properties of GLS. While this is different from the proposed Bayesian method where the covariance matrix is unknown, these properties will give intuition on the relationship between the covariance parameters and inference on the regression coefficient.

4.2.1 Generalized Least Squares

Let $C \in \mathbb{R}^{n\times n}$ be a positive semi-definite matrix of rank r for $r \leq n$. Let $C = P\Lambda P^T$ denote the principal component or eigen-decomposition of C, where $P \in \mathbb{R}^{n\times r}$ denotes the column-matrix of principle component vectors and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_r)$ denotes the diagonal matrix of non-zero eigenvalues. Let $C^+ \equiv P^T\Lambda^{-1}P$ denote the Moore-Penrose pseudo-inverse of C. Assuming that all components of C and the vectors $\mathbf{x}$ and $\mathbf{y}$ are known, under the GLS formulation of Equation (4.1) the maximum likelihood estimator of β is

$$\hat{\beta} = \frac{\mathbf{x}^T C^+ \mathbf{y}}{\mathbf{x}^T C^+ \mathbf{x}}. \tag{4.2}$$

This will be proved in the proceeding sub-section. It should be noted that when C is full rank, or $r = n$, then $C^{-1} \equiv C^+$ can be substituted into the relevant equations.

Denote the projection of $\mathbf{x}$ and $\mathbf{y}$ onto the eigenspace of C as

$$\mathbf{x}^* = P^T\mathbf{x} \text{ and } \mathbf{y}^* = P^T\mathbf{y}.$$

Then, the numerator of Equation (4.2) becomes $\mathbf{x}^T C^+ \mathbf{y} = \mathbf{x}^T P\Lambda^{-1}P^T\mathbf{y} = (\Lambda^{-1/2}\mathbf{x}^*)^T(\Lambda^{-1/2}\mathbf{y}^*)$ and the denominator becomes $\mathbf{x}^T C^{-1}\mathbf{x} = \mathbf{x}^T P\Lambda^{-1}P^T\mathbf{x} = (\Lambda^{-1/2}\mathbf{x}^*)^T(\Lambda^{-1/2}\mathbf{x}^*)$. So, writing out these inner products as summations it can be seen that the MLE of β is

$$\hat{\beta} = \frac{\sum_{i=1}^r x_i^* y_i^*/\lambda_i}{\sum_{i=1}^r (x_i^*)^2/\lambda_i} \tag{4.3}$$

and it can be seen (see Section 4.2.2) that the standard error is

$$se(\hat{\beta}) = \left(\sum_{i=1}^{r} \left(x_i^*\right)^2 / \lambda_i \right)^{-1/2}. \tag{4.4}$$

With these equations, the GLS formulation can be viewed as ordinary least squares with heteroskedastic errors performed on the patterns projected onto the eigenspace of C. In particular, Equation (4.1) can be rewritten as

$$\mathbf{y}^* \sim \beta \mathbf{x}^* + \epsilon;\ \forall_i \epsilon_i \sim N(0, \lambda_i). \tag{4.5}$$

When C is rank deficient, then $\mathbf{x}^*, \mathbf{y}^* \in \mathbb{R}^r$, and the transformation implied by the basis vector matrix P is a projection onto a lower-dimensional subspace. When C is full rank, this projection can be seen as a rotation in a direction where the coordinates of the rotated vector are statistically independent.

In practice, the covariance matrix C is not known and must be estimated. The empirical covariance matrix for the pre-industrial trend fields is

$$\hat{C} = \frac{1}{n_{\mathcal{P}_c}} \sum_{k=1}^{n_{\mathcal{P}_c}} (\mathbf{z}_k - \overline{\mathbf{z}})^T (\mathbf{z}_k - \overline{\mathbf{z}}). \tag{4.6}$$

The rank of this matrix will be equal to the number of control vectors $n_{\mathcal{P}_c}$, which is generally much less than n_{grid}; in the case considered here, $n_{\text{grid}} = 2592$ and the largest value of $n_{\mathcal{P}_c}$ is 66 (Table 4.1). Let $\hat{\lambda}_1, \ldots, \hat{\lambda}_{n_{\text{grid}}}$ denote the eigenvalues of the empirical covariance matrix $\hat{C}$. When $n_{\mathcal{P}_c} < n_{\text{grid}}$, $\hat{\lambda}_i = 0$ for all $i > n_{\mathcal{P}_c}$. If the true covariance matrix C is full-rank, it will have no zero eigenvalues. Ignoring for the moment that the eigenvectors estimated from $\hat{C}$ will be different from the eigenvectors of the true covariance matrix C, the values $\hat{\lambda}_i$ will be under-estimates of corresponding true non-zero eigenvalues. When $\hat{C}$ is treated as the true covariance matrix, these zero eigenvalues do not pose a mathematical problem for the GLS estimation in Equation (4.3) as the dimensions orthogonal to the estimated principal components are effectively removed in the subspace projection. However, $\hat{\lambda}_i$ will be under-estimated but non-zero for higher values of i which are nevertheless less than $n_{\mathcal{P}_c}$,

which will affect equations (4.3) and (4.4). The fact that the higher-number eigenvectors will themselves be inaccurately estimated adds further reason that the inclusion of those components may yield inaccurate results. This issue will be discussed in further depth in Section 4.2.6, but motivates the truncated principal component parameterization of the covariance matrix, which in essence projects the data onto a lower-dimensional subspace than implied by the rank of the empirical covariance matrix in order to avoid those poorly estimated components from impactings the reliability of the regression results.

4.2.2 Proof of GLS MLE and Standard Error

This section will briefly sketch the proofs the maximum likelihood estimator (MLE) of the GLS method and the formula for the standard error:

Proof of Equation (4.2)*:* Let $C \in \mathbb{R}^{n \times n}$ be a positive semi-definite matrix of rank $r \leq n$. Let $\mathbf{y} = \beta\mathbf{x} + \epsilon$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\epsilon \sim N(0, C)$. Then the MLE estimator for β of $\mathbf{y} = \beta\mathbf{x} + \epsilon$ is $\hat{\beta} = \frac{\mathbf{x}^T C^+ \mathbf{y}}{\mathbf{x} C^+ \mathbf{x}}$, where C^+ is the Moore-Penrose psuedo-inverse defined as $C^+ = P^T \Lambda^{-1} P$ where B_r is the $r \times n$ matrix of eigenvectors and $\Lambda_r = \text{diag}(\lambda_1, \ldots, \lambda_r)$ is the diagonal matrix of eigen values of C. Under the Lebesgue measure [Rao, 1973] the density function for a rank-deficient multivariate normal distribution yields a log-likelihood function of

$$\ell(\beta) = c - \frac{1}{2}(\mathbf{y} - \beta\mathbf{x})^T C^+ (\mathbf{y} - \beta\mathbf{x})$$

where c is a constant not depending on β. Let $\mathbf{x}^* = P^T\mathbf{x}$ and $\mathbf{y}^* = P^T\mathbf{y}$. Then the likleihood is equivalent to

$$\ell(\beta) = c - \frac{1}{2}(\mathbf{y}^* - \beta\mathbf{x}^*)^T \Lambda^{-1} (\mathbf{y}^* - \beta\mathbf{x}^*)$$

Taking the derivative with respect to β yields

$$\frac{\partial}{\partial\beta}\ell(\beta) = c' - \beta(\mathbf{x}^*)^T\Lambda^{-1}\mathbf{x}^* + (\mathbf{x}^*)^T\Lambda^{-1}\mathbf{y}^*$$

whose solution is

$$\hat{\beta} = \frac{(\mathbf{x}^*)^T\Lambda^{-1}\mathbf{y}^*}{(\mathbf{x}^*)^T\Lambda^{-1}\mathbf{x}^*} = \frac{\mathbf{x}^T C^+ \mathbf{y}}{\mathbf{x}^T C^+ \mathbf{x}}$$

Proof of Equation (4.4)*:* Taking the variance of $\hat{\beta}$ yields

$$\text{var}[\hat{\beta}] = \frac{\mathbf{x}^T C^+ \text{var}[y] C^+ \mathbf{x}}{(\mathbf{x}^T C^+ \mathbf{x})} = \frac{\mathbf{x}^T C^+ C C^+ \mathbf{x}}{(\mathbf{x}^T C^+ \mathbf{x})^2} = \frac{\mathbf{x}^T C^+ \mathbf{x}}{(\mathbf{x}^T C^+ \mathbf{x})^2} = (\mathbf{x}^T C + \mathbf{x})^{-1}$$

$$= \left(\sum_{i=1}^{r} (x_i^*)^2 / \lambda_i \right)^{-1/2}$$

4.2.3 Truncated principal Component Parameterization

The eigen-decomposition $C = P \Lambda P^T$ can be equivalently written as $C = \sum_{i=1}^{r} \lambda_i p_i p_i^T$, where p_i denote the eigenvector columns of P. The idea behind the truncated eigenvector representation is to choose an integer $1 \leq \kappa \leq n_{\mathcal{P}_c}$ and define the rank-κ covariance matrix

$$C_{\kappa;\text{PC}} \equiv \sum_{i=1}^{\kappa} \lambda_i p_i p_i^T \equiv P_\kappa \Lambda_\kappa P_\kappa^T \tag{4.7}$$

where P_κ is the column matrix of the first κ principal components and $\Lambda_\kappa = \text{diag}(\lambda_1, \ldots, \lambda_\kappa)$. Here C is given the subscript "PC" to differentiate it from the Laplacian parameterization that will be introduced in the next section. This covariance matrix can then be used with Equation (4.2) and (4.4) to obtain GLS results in the frequentist setting. Traditionally, the principle component vectors and the integer κ are selected and treated as fixed prior to inference, and the uncertainty in their estimation are is not propagated to the final conclusions. The remainder of this sub-section will discuss issues with treating the principal component vectors p_i as fixed, with issues in treating κ as fixed to be discussed in Section 4.2.6.

One issue with treating the principal component vectors are fixed is that these components will contain estimation error due to the limited number of available control vectors. To visualize this error, 13 trend fields were randomly sub-selected from the 66 trend fields of the GISS-E2-1-G model. Then the first principal component was computed from the sub-sampled vectors as well as the difference between this component and the first principal component obtained from using all 66 control vectors. This was repeated for nine random

sub-samples with the results displayed in Figure 4.1. These patterns can be interpreted as the error from estimating the first principal component using 13 control vectors if the covariance matrix implied by the full control vector set is treated as the "truth". The GISS-E2-1-G model was chosen as it has the highest number of pre-industrial control vectors and the number 13 was chosen to correspond to AWI-CM-1-1-MR which has the smallest number of control vectors (Table 4.1). For reference, the first principal component for the "true" covariance of the GISS-E2-1-G model is displayed in the second column and third row of Figure 4.2. It can be seen in Figure 4.1 that even for the first principal component there is notable error caused by the limited number of control runs. This visualization gives an idea of the scale and patterns of the error when estimating principal components from a limited number of control vectors; in reality, the true error for this number of components in this experiment would be larger due to the error between the covariance matrix implied by the 66 control vectors and the true covariance matrix.

A second issue with the use of principal components is that they rely on the particular climate model used, and the covariance matrices for different climate models may differ from each other and from the "true" covariance structure. In particular, the "true" covariance matrix generating the field of observations can be imagined as being as different from the covariance matrices implied by the climate models as they are from each other. Under this conceptual framework, it can be seen that there may be a considerable "mismatch" between the components of the true, unobserved covariance matrix and those estimated using climate models, even when the number of principal components is restricted to the lower-number better-estimated patterns. Furthermore, in the inference procedure, these components are treated as fixed in advance, and there has not been any way developed to incorporate the uncertainty in their estimation into the optimal fingerprinting procedure. To get a sense of the difference between the covariance structures implied by the different climate models, Figure 4.2 shows the estimated first principal components for each of the 16 models under consideration here (Table 4.1). This shows notable heterogeneity between the

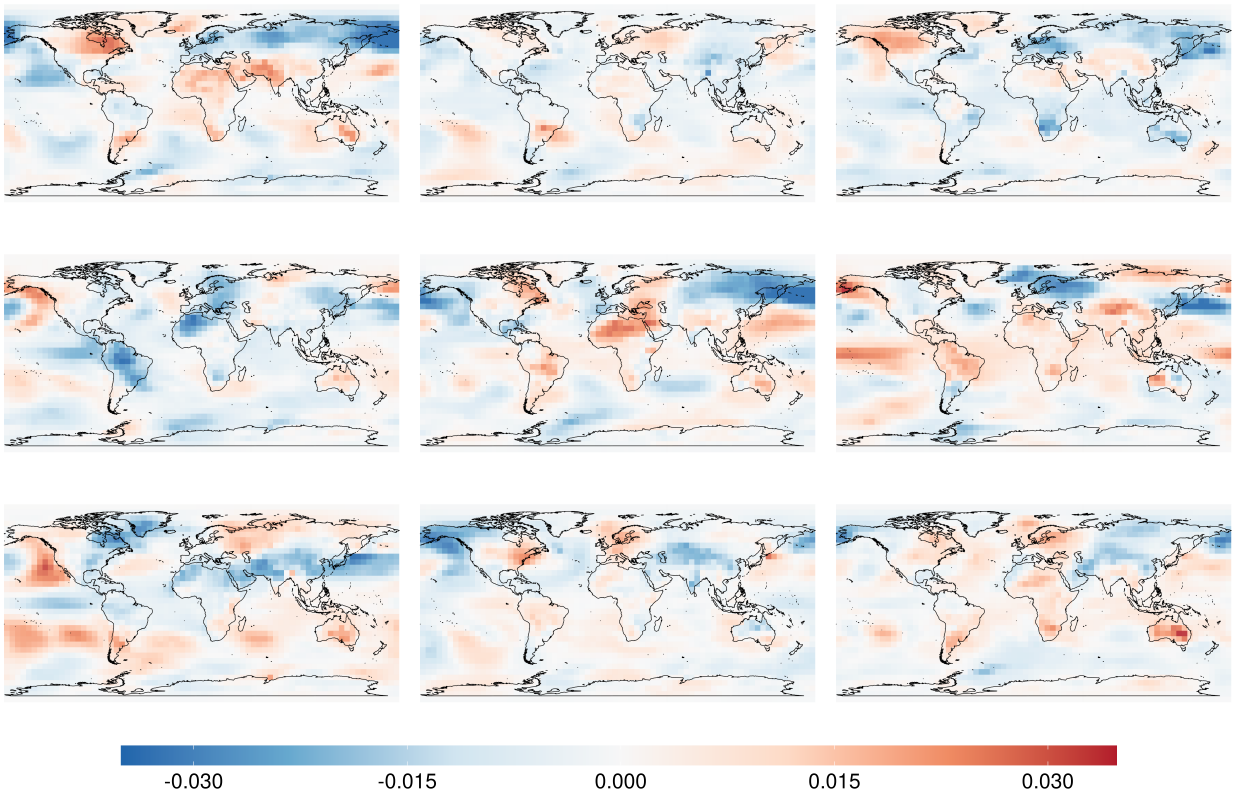


Figure 4.1: Error in estimating the first principal component of the GISS-E2-1-G pre-industrial covariance matrix when only 13 control vectors are used. Displayed values are the difference between the principal components of nine random sub-samplings of the control vectors minus the first principal component when all 66 vectors are used. The number 13 was chosen to correspond to the minimum value of $n_{\mathcal{P}_c}$ in Table 4.1. For comparison, the fully sampled principal component can be seen in Figure 4.2.

patterns expressed in the first principal components of these different covariance matrices.

The dependence of the principal component vectors on the size and properties of the climate models used presents a source of variability and uncertainty to the optimal fingerprinting procedure which is not easily represented. A third issue with the use of principal components is that they are not guaranteed to capture spatially coherent patterns. The first principal components visualized in Figure 4.2 appear to capture spatially coherent coherent patterns, however this is a consequence of correlations present in the data as opposed to a property of the method. While it generally appears to be the case that lower-number principal components capture meaningful patterns, higher principal components are more likely to correspond to patterns that do not exhibit smooth spatial variation. Given the largely spatial nature of the fields under consideration in optimal fingerprinting, patterns that do not exhibit smooth spatial variation are likely to represent "noise" rather than meaningful modes of variability. This highlights the desirability of a basis function parameterization which enforces spatial correlation.

It may be asked at this point why it is undesirable to have "noise" patterns represented in the covariance matrix for the residual term, given the fact that residual terms are usually interpreted as "noise" or error from the GLS perspective. The reason is that the residual term in (4.1) does not actually represent "noise" or error, but rather represents natural variability, which is an integral aspect of the climate system. As mentioned in Section 4.2.1, in the case where the covariance matrix is full-rank, GLS can be viewed as a heteroskedastic linear regression performed on the vectors projected onto the principal components. In the case that the covariance matrix is rank-deficient, this projection is onto a lower-dimensional subspace. In either case, if basis vectors corresponding to "noise" are included, then the regression will take those components in the data into account for both the regressors and the response. This would be undesirable, as then the estimate of β will be based on information from components that do not represent meaningful patterns from a detection and attribution perspective. This would presumably lead to results that do not optimally recover the true

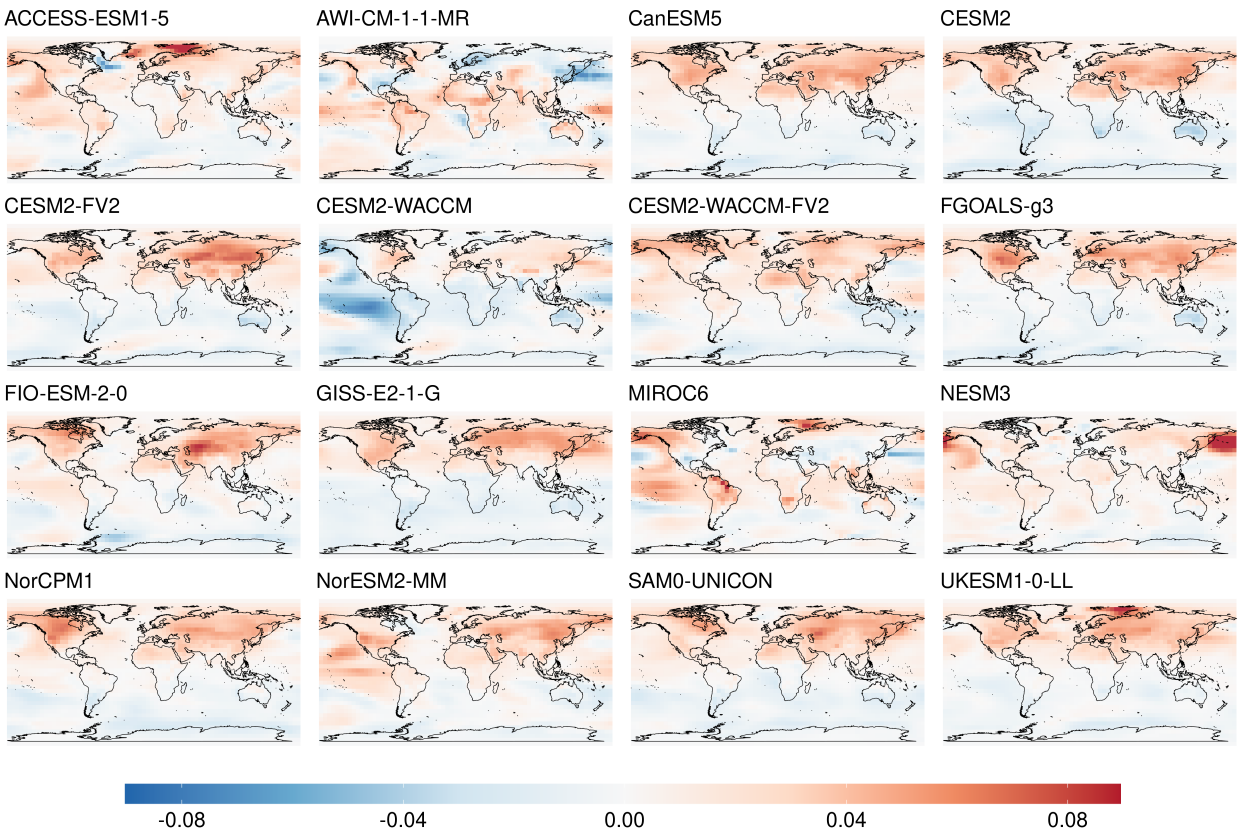


Figure 4.2: First principal components, or EOFs, for the 16 pre-industrial climate ensembles used. Components were calculated using area-weighted fields of temperature trends over 1990-2015.

value, a presumption which will be empirically verified in the validation study of Section 4.3.

Limited data contributes to this phenomenon in the principal component case since it makes it more likely that "noise" components are included and assigned unnaturally high precision values, however, this is not fundamentally an issue of limited data. In the case where the true full-rank covariance matrix was known exactly, it would be expected that many or most of the principal components would correspond to "noise" patterns, as in a full-rank matrix the eigenvectors will by definition span the space of all possible patterns. In other words, the number of degrees of freedom in the trend fields will be expected to be much smaller than the grid-cell dimension n_{grid}. The implications of this issue for inference can be seen in Equation (4.2), where β is driven by the correspondence between y_i and x_i for all of the included components. If these components represent "noise", such influence on β would be unwarranted. This issue becomes more severe when component variances are under-estimated, as is generally the case for higher-number components with limited data, as the "noise" components in equation (4.2) will be scaled upward to a higher degree than would be warranted by the true variability structure. This under-estimation issue could also be present with sufficient information if the estimated covariance matrix diverges from the "true" covariance matrix due to model error, which would be present even with sufficient control data.

4.2.4 Laplacian Parameterization

Given the issues with the principal component parameterization, it would be desirable to have a covariance matrix parameterization that relies on basis functions that do not depend on data. Ideally, these basis functions would be general enough to capture a wide potential variety of covariance patterns and would enforce spatial correlation in order to prioritize the inclusion of meaningful patterns. The eigenfunctions of the Laplace operator on the sphere yield basis vectors with these properties [Courant and Hilbert, 2008, DelSole and Tippett,

2015]. Specifically, the Laplace operator on the sphere is defined as

$$\nabla^2 Y = \frac{1}{\cos^2(\text{lat})}\left[\frac{\partial^2 Y}{\partial \text{lon}^2} + \cos(\text{lat})\frac{\partial}{\partial \text{lat}}\frac{\partial Y}{\partial \text{lat}}\right]$$

on latitudinal and longitudinal coordinates. Let $(\text{lat}_i, \text{lon}_i)$ denote the latitudinal and longitudinal coordinates of the i-th grid-cell for $1 \leq i \leq n_{\text{grid}}$. Assuming a grid with equally spaced latitudinal and longitudinal dimensions, let δlon denote the distance between points in the longitudinal dimension and let δlat denote the distance between points in the latitudinal dimension. Let $d_{gc}(i,j)$ denote the great circle distance between the i-th and j-th gridpoints. Then the discretized Laplacian operator can be represented by the matrix $A \in \mathbb{R}^{n_{\text{grid}} \times n_{\text{grid}}}$ with elements

$$A_{ij} = \begin{cases} -\frac{4}{\pi}\log(2(\sin(d_{gc}(i,j)^2)))\,\delta\text{lat}\,\delta\text{dlon}\,\sqrt{\cos(\text{lat}_i)\,\cos(\text{lat}_j)} & \text{for } i \neq j \\ \frac{1}{4}\rho_i^2\left(1 - 2\log\frac{\rho_i}{\sqrt{2}}\right) & \text{for } i = j \end{cases}$$

where

$$\rho_i = \sqrt{\delta\text{lat}\,\delta\text{lon}\cos(\text{lat}_i)/\pi}$$

is the area of a circle whose area approximates the area of the i-th grid-cell.

Let A^* denote A projected onto the orthogonal component of the constant vector; this is done in order to ensure that the constant vector is an eigenfunction of the discretized Laplace operator. Let $\ell_1, \ldots, \ell_{n_{\text{grid}}}$ denote the normalized eigenvectors of the matrix A^* where the constant pattern is re-ordered to come first. These vectors will be referred to as "Laplacian basis vectors". The reader is directed to DelSole and Tippett [2015] for further details on their construction. To get a sense for the spatial relationships captured by these functions, the first nine eigenvectors on the sphere are displayed in Figure 4.3. Here it can be seen that these basis vectors capture spatially coherent correlation patterns at scales that decrease as the basis vector number increases.

Let $L \in \mathbb{R}^{n_{\text{grid}}, n_{\text{grid}}}$ denote the column vector matrix of $\ell_1, \ldots, \ell_{n_{\text{grid}}}$. Define the Laplace-parameterized covariance matrix as $C_{\text{Laplace}} = L\Lambda_{\text{Laplace}}L^T = \sum_{i=1}^{n_{\text{grid}}} \lambda_{i;\text{Laplace}}\ell_i\ell_i^T$ where

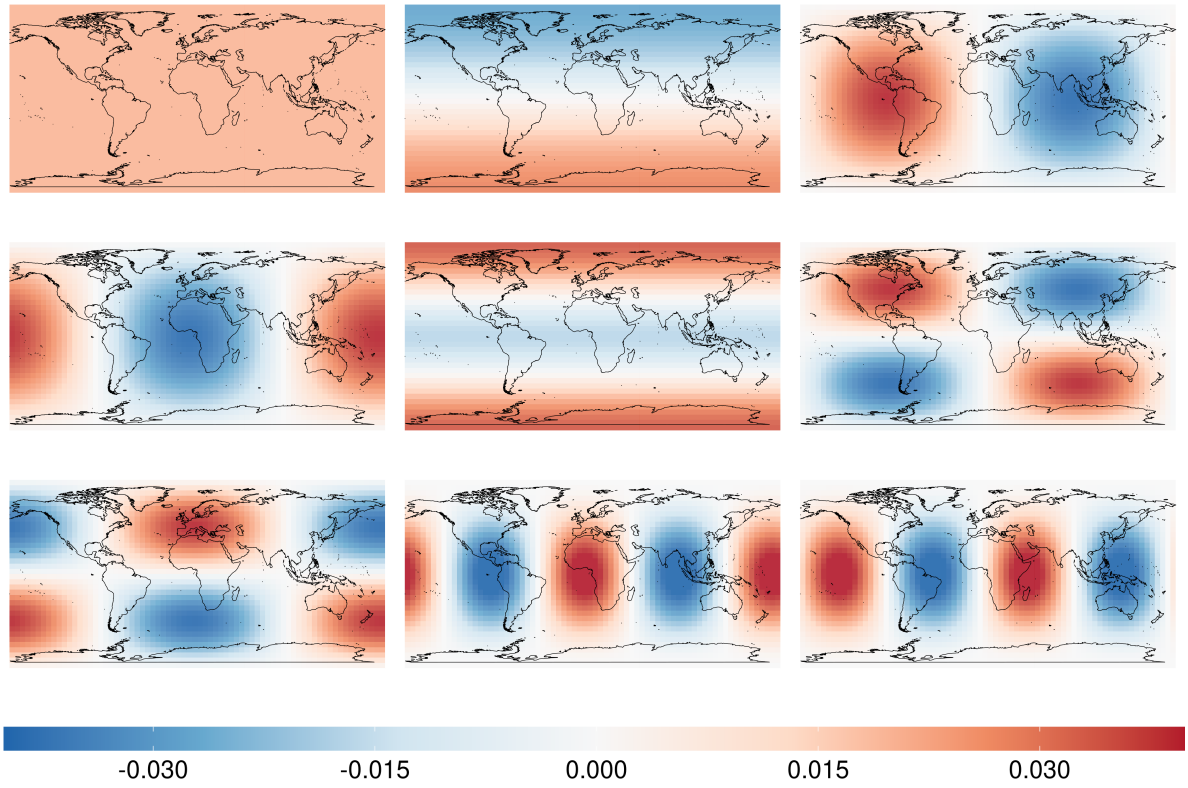


Figure 4.3: First nine Laplacian basis functions on the sphere.

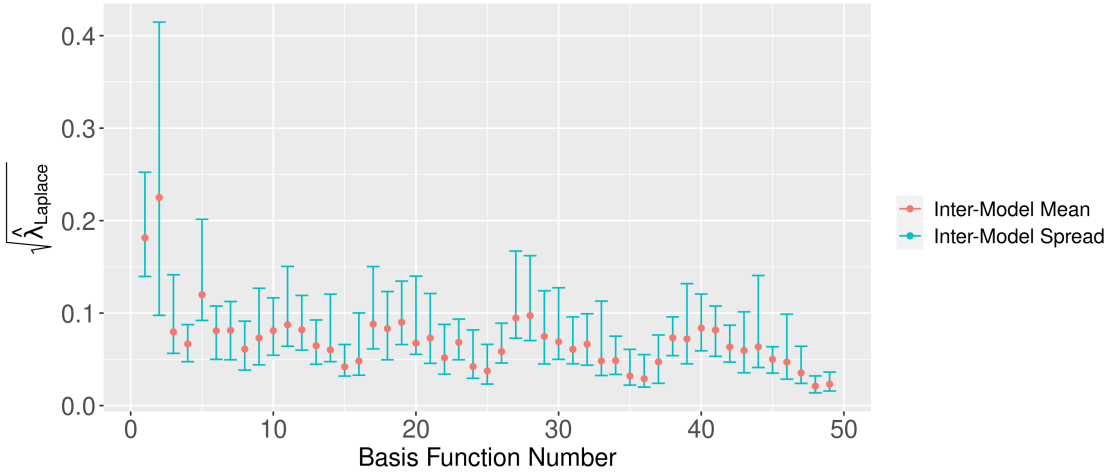


Figure 4.4: Empirical standard deviations for the first fifty Laplacian basis functions. Inter-model means for each coefficient are over the 16 pre-industrial control models of Table 4.1. Intervals show the maximum and minimum values of the coefficients over the 16 models.

$\Lambda_{\text{Laplace}} = \text{diag}(\lambda_{1;\text{Laplace}}, \ldots, \lambda_{n_{\text{grid}};\text{Laplace}})$ is a diagonal matrix of variance coefficients for the Laplace basis vectors. The truncated Laplace covariance matrix is then defined as

$$C_{\kappa;\text{Laplace}} \equiv \sum_{i=1}^{\kappa} \lambda_{i;\text{Laplace}} \ell_i \ell_i^T \equiv L_\kappa \Lambda_{\kappa;\text{Laplace}} L_\kappa^T \tag{4.8}$$

where κ is the truncated number of Laplace components, L_κ is the column vector matrix of the first κ Laplace eigenfunctions, and $\Lambda_{\kappa;\text{Laplace}} = \text{diag}(\lambda_{1;\text{Laplace}}, \ldots, \lambda_{\kappa;\text{Laplace}})$.

As before, let $\{\mathbf{z}_1, \ldots, \mathbf{z}_{n_{\mathcal{P}_c}}\}$ denote a collection of control-run vectors for pre-industrial climate simulation $\mathcal{P}_c$. Let $\mathbf{z}_k^* = L^T \mathbf{z}$. Then the empirical Lapalacian variance coefficients are

$$\hat{\Lambda}_{\text{Laplace}} = \text{diag}\left(\frac{1}{n_{\mathcal{P}_c}} \sum_{k=1}^{n_{\mathcal{P}_c}} \left(\mathbf{z}_k^* - \overline{\mathbf{z}^*}\right)^T \left(\mathbf{z}_k^* - \overline{\mathbf{z}^*}\right) \right) \text{ where } \mathbf{z}_k^* = L^T \mathbf{z} \tag{4.9}$$

and $\overline{\mathbf{z}^*} = \frac{1}{n_{\mathcal{P}_c}} \sum_{k=1}^{n_{\mathcal{P}_c}} \mathbf{z}_k^*$. The diagonal entries of $\hat{\Lambda}$ will be denoted as $\hat{\Lambda}_{\text{Laplace}} = (\hat{\lambda}_{1;\text{Laplace}}, \ldots, \hat{\lambda}_{n_{\text{grid}};\text{Laplace}})$. This construction is analagous to the principal component case if L is replaced with P. To see this, left-multiplying Equation (4.6) by P^T and then right-multiplying the result by P yields $P^T \hat{C} P = \Lambda = \left(\frac{1}{n_{\mathcal{P}_c}} \sum_{k=1}^{n_{\mathcal{P}_c}} \left(\mathbf{z}_k^* - \overline{\mathbf{z}^*}\right)^T \left(\mathbf{z}_k^* - \overline{\mathbf{z}^*}\right) \right)$ which is a diagonal matrix by construction.

To get a sense for the values of the Laplacian variance coefficients, Figure 4.4 displays the values of $\sqrt{\hat{\lambda}_i}$, which represent the standard deviation for the i-th Laplacian component, for $1 \leq i \leq 50$. For each component, means and max-min spreads calculated over the 16 pre-industrial control models are displayed. It can be seen that there remains a noticeable degree of inter-model heterogeneity; although unlike with the principal components displayed in Figure 4.2 these differences can be visualized succinctly due to the fact that the underlying basis function representations are the same between models. There also appears to be a periodic oscillation in the Laplacian coefficients. This is due to the fact that the Laplacian basis vectors oscillate between primarily longitudinal correlation patterns and primarily latitudinal correlations as can be seen in Figure 4.3. The former correlation patterns are more prominent in the data since temperature changes are more correlated across longitude than across latitude.

4.2.5 Bayesian Hierarchical Model

In the previous two sections P has denoted the column matrix of principal component basis vectors and L has denoted the matrix of Laplacian basis vectors. For the remainder of this section B will denote a generic set of basis functions standing for either principal components or Laplacian basis vectors depending on which method is used. Let $\mathcal{K}$ denote a random variable for the number of components with discrete probability distribution $P(\mathcal{K} = \kappa)$ over $1 \leq \kappa \leq n_{\text{basis}}$; here $n_{\text{basis}} = \min(n_{\mathcal{P}_c}, n_{\text{grid}})$ in the principal component case and $n_{\text{basis}} = n_{\text{grid}}$ in the Laplacian case. Conditioning on $\mathcal{K} = \kappa$, let $C_\kappa = B_\kappa^T \Lambda_\kappa B_\kappa$ represent the truncated covariance matrix with generic basis vectors and coefficients $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_\kappa)$. Let $\hat{\lambda}_1, \ldots, \hat{\lambda}_{n_{\text{grid}}}$ refer to either the eigenvalues of the empirical covariance matrix $\hat{C}$ defined in Equation (4.6) when the principal component basis is used or the empirical Laplacian coefficients defined in Equation (4.9) when the Laplacian basis is used. As in Section 4.2.1, let $\mathbf{y}^* = B^T\mathbf{y}$ and let $\mathbf{x}^* = B^T\mathbf{x}$.

Conditional on $\mathcal{K} = \kappa$, the Bayesian optimal fingerpriting model can be defined concisely

with the following equations:

$$y_i^* - \beta x_i^* \sim N(0, \lambda_i) \text{ for } 1 < i \leq \kappa \tag{4.10}$$

$$\log \lambda_i \sim N(\log \hat{\lambda}_i, 1) \text{ for } 1 < i \leq \kappa$$

$$\beta \sim \text{Unif}(-\text{Inf}, \text{Inf})$$

where regression parameter β is given a dispersed, uninformative prior in order to avoid biasing the results towards a particular value.

This basic framework is similar to and inspired by the approach of KHS17. One difference is that the above ignores the fact that $\mathbf{y}$ and $\mathbf{x}$ come from ensembles and rather take them to be fixed at the ensemble means. This simplification with regards to the forced ensemble would not be expected to affect the results, since the approach of KHS17 integrates the forced variability out of the likelihood function and as such it does not influence the posterior distributions. Observational uncertainty is likely to be small in the field of 25-year trends over the time period under consideration. As such this uncertainty is omitted here, although in principle the likelihood techniques of KHS17 could be used to represent observational uncertainty in a computationally efficient way.

Another difference between the framework outlined above and that of KHS17 is that their approach introduced a parameter $\sigma^2 > 0$ to be added to the diagonal of the covariance matrix. Conditioning on $\mathcal{K} = \kappa$, this yields the full-rank $C_{\kappa;\text{fr}} = B_\kappa^T \Lambda_\kappa B_\kappa + \sigma^2 I_{n_{\text{grid}}}$, where the subscript "fr" stands for full-rank, as opposed to the potentially rank-deficient definition. The rank deficient approach allows for the "noise" components to be eliminated from the regression. Specifically, say that $\kappa < n_{\text{grid}}$ and B is the matrix of eigenvectors of C. The rank-deficient covariance matrix can be written as $C_\kappa = B_\kappa^T \Lambda_\kappa B_\kappa \equiv B_{n_{\text{grid}}}^T \Lambda_{n_{\text{grid}}} B_{n_{\text{grid}}}$ where $B_{n_{\text{grid}}}$ is equal to B with the addition of $n_{\text{grid}} - \kappa$ vectors spanning the orthogonal complement of the κ eigenvectors and $\Lambda_{n_{\text{grid}}} = \text{diag}(\lambda_1, \ldots, \lambda_\kappa, 0, \ldots, 0)$. When σ^2 is added to the diagonal, this becomes $C_{\kappa;\text{fr}} = B_\kappa^T \Lambda_\kappa B_\kappa + \sigma^2 I_{n_{\text{grid}}} = B_{n_{\text{grid}}}^T \Lambda_{\kappa;\text{fr}} B_{n_{\text{grid}}}$ where $\Lambda_{\kappa;\text{fr}} = \text{diag}(\lambda_1 + \sigma^2, \ldots, \lambda_{n_{\text{pi}}} + \sigma^2, \sigma^2, \ldots, \sigma^2)$. Then, returning to the formulas in Section 4.2.1, the regression equation

$\mathbf{y}^* - \beta\mathbf{x}^* \sim N(0, \Lambda_{\kappa;\text{fr}})$ holds with $\mathbf{x}^* = \Lambda^{-1/2} B_{n_{\text{grid}}}^T \mathbf{x}$, where $\mathbf{x}^*$ is a vector of length n_{grid} rather than κ. The value of κ then determines whether or not the variance for the i-th component is $\lambda_i + \sigma^2$ or σ^2; this contrasts from the rank-deficient version where the i-th component is removed from the regression if $i > \kappa$. In other words, in this version κ no longer has an interpretation of selecting the number of components as all of the components will be used. It is desirable to limit the number of components for the reasons discussed at the end of Section 4.2.3 and justified with empirical results in Section 4.3.3.

While using a rank-deficient covariance matrix in the frequentist context yields the closed-form estimators presented in Section 4.2.1, there is a slight complication from a Bayesian perspective as the likelihood for $y_i - \beta x_i$ in Equation (4.10) is not defined if $i > \kappa$. A way to get around this is to let $\mathbf{1}\{i \leq \kappa\}$ denote the indicator function for $i \leq \kappa$ and write Equation (4.10) as

$$(y_i^* - \mathbf{1}\{i \leq \kappa\}\beta_\kappa x_i^*) \sim N(0, \lambda_{i;\kappa}) \text{ for } 1 \leq i \leq n_{\text{grid}} \tag{4.11}$$

where κ is added as a subscript to β and λ_i in order to represent the dependency between these parameters.

4.2.6 Selecting Covariance Truncation Number

Selection of the number of basis functions in the covariance matrix representation has long been identified as an important issue whose choice significantly impacts the results of optimal fingerprinting [Hegerl et al., 1996]. Concerns have primarily been driven by the accuracy of under-estimating component variances, however as discussed at the end of Section 4.2.3 when the principal component basis is used there is also concern about estimating the principal components themselves. In either case, however, inaccurate estimates of the component variances can lead to inaccurate optimal fingerprinting results and confidence intervals. Recalling the frequentist MLE $\hat{\beta} = \frac{\sum_i x_i^* y_i^* / \lambda_i}{\sum_i (x_i^*)^2 / \lambda_i}$, if λ_i is unrealistically small, then the i-th components will be given larger weights thus will have a disproportionate impact on the estimate of β.

Furthermore, recall from the standard error formula $se(\hat{\beta}) = \left(\sum_i (x_i^*)^2/\lambda_i\right)^{-1/2}$ that smaller values of λ_i will increase the value of $(x_i^*)^2/\lambda_i$. This will decrease the standard error $se(\hat{\beta})$ and lead to over-confident estimates. Conversely, if λ_i is over-estimated, then it could under-weight important components and yield less precise detection and attribution results and overly conservative error bars. As mentioned at the end of Section 4.2.1, under-estimation of component variances is virtually guaranteed to happen in the principal component representation of the covariance matrix when $n_{\mathcal{P}_c} < n_{\text{basis}}$. While n_{grid} Laplacian basis functions patterns are defined regardless of the value of $n_{\mathcal{P}_c}$, higher number Laplacians will exhibit increasingly small correlation lengths and as such would not likely correspond to meaningful patterns. This will lead to variances which are poorly estimated from the control vectors, and as such the Laplacian approach is not immune to the model selection issue. While the properties of the frequentist estimators were used to drive the intuition presented here, the logic largely transfers to the Bayesian setting as well.

Within the frequentist framework the most common method employed for selecting the number of components is the "residual consistency" test originally proposed by [Allen and Tett, 1999]. This test is also used in the Bayesian framework of KHS17 for post-hoc model validation. The motivation for the test relies on the observation that when the covariance structure is being modeled accurately, the squared scaled and projected residuals will be χ^2 distributed. To see this, recall that $\mathbf{y}^* = B^T\mathbf{y}$ and $\mathbf{x}^* = B^T\mathbf{x}$ are the observed and forced vectors respectively projected onto the basis functions implied by the covariance matrix parameterization. Let $r_{\kappa;i} = \frac{y_i^* - \beta x_i^*}{\sqrt{\lambda_i}}$ denote the normalized residual conditioned on a value of β. It can be seen from Equation (4.10) that

$$r_{\kappa;i} \equiv \frac{y_i^* - \beta x_i^*}{\sqrt{\lambda_i}} \sim N(0,1).$$

Define the vector of normalized residuals as $\mathbf{r}_\kappa = [r_{\kappa;1}, \ldots, r_{\kappa;\kappa}] = \Lambda_\kappa^{-1/2} B_\kappa^T(\mathbf{y} - \beta\mathbf{x})$. It follows that, conditioned on β, $(r_{\kappa;i})^2 \sim \chi_1^2$, and that

$$\mathbf{r}_\kappa^T \mathbf{r}_\kappa = \sum_{i=1}^{\kappa} (r_{\kappa;i})^2 = \sum_{i=1}^{\kappa} \frac{(y_i^* - \beta x_i^*)^2}{\lambda_i} \sim \chi_k^2 \tag{4.12}$$

When not conditioning on β, the degrees of freedom are reduced by one, yielding $\mathbf{r}_\kappa^T \mathbf{r}_\kappa = (\mathbf{y} - \beta\mathbf{x}) C_\kappa^{-1} (\mathbf{y} - \beta\mathbf{x}) \sim \chi_{\kappa-1}^2$ which is the form of the test stated in Equation 18 of Allen and Tett [1999] when the number of forcing patterns is equal to one.

The residual consistency test is then rejected for a particular κ if $\mathbf{r}_\kappa^T \mathbf{r}_\kappa$ is larger than an α-significance threshold with respect to the χ^2 distribution. Looking at the summands of Equation (4.12), it can be seen that large values will be achieved when the component variances λ_i are small relative to the squared residuals $(y_i^* - \beta x_i^*)^2$. As such, this test is more likely to reject the number of components if the values of λ_i under-estimate the true variance. The motivation for the test rejecting only high values of the test statistic is to avoid misleading and over-confident conclusions, however in principle a two-sided test could be used to ensure that the component variances are not over-estimated as well. The number of components is generally obtained by evaluating the results of the test for increasing values of κ until the test is rejected, at which point κ is fixed at the previous passing value.

While useful, the residual consistency test carries significant drawbacks. One drawback is that the significance level α is somewhat arbitrary. A more conceptual difficulty is that in the hypothesis testing framework there is a difference between failing to reject a test and the null hypothesis being true. In the case of the residual consistency test, the null hypothesis is that the squared scaled and de-correlated residuals are χ^2 distributed, which can be viewed as saying that the covariance matrix structure is accurate. A rejection of this test means that there is sufficient evidence from the observations to show that the estimated covariance structure is inadequate. Failure to reject the test means that there is not sufficient evidence to reject, but does not mean that the covariance structure is itself accurate. A third drawback of this procedure is that it is generally done prior to and separately from inference, and as such the uncertainty in choosing the number of components in this way is not incorporated into the model. Finally, concern has been raised that the

statistical assumptions underlying the validity of the test in a frequentist framework may not be satisfied in practice [McKitrick, 2022]. Despite these drawbacks, the residual consistency test remains a well-motivated approach to choosing the number of components to minimize underestimation of the component variances.

A Bayesian approach that incorporates component selection within the inference procedure has the potential to remedy these drawbacks. This is the motivation for the framework proposed by KHS17, who estimate the distribution over the number of components by fitting the regression model separately for each value of $\mathcal{K} = \kappa$. With these fitted models, Bayesian model averaging [Hoeting et al., 1999] is used to marginalize over $\mathcal{K}$ and obtain a posterior distribution for β that takes the uncertainty in $\mathcal{K}$ into account. More specifically, the posterior distribution of β conditioned on the number of components can be written as $[\beta|\mathcal{K} = \kappa, \mathbf{y}, \mathbf{x}]$. Once these posterior distributions are obtained through MCMC, the marginal posterior of β is

$$[\beta|\mathbf{y}, \mathbf{x}] = \sum_{k=1}^{n_{\text{basis}}} [\beta|\mathcal{K} = \kappa, \mathbf{y}, \mathbf{x}] P(\mathcal{K} = \kappa|\mathbf{y}, \mathbf{x}) \tag{4.13}$$

where

$$P(\mathcal{K} = \kappa|\mathbf{y}, \mathbf{x}) \propto [\mathbf{y}, \mathbf{x}|\mathcal{K} = \kappa] P(\mathcal{K} = \kappa) \propto [\mathbf{y}, \mathbf{x}|\mathcal{K} = \kappa]$$

assuming a flat prior for $P(\mathcal{K} = \kappa)$. The Bayesian modeling averaging formula of Hoeting et al. [1999] yields

$$[\mathbf{y}, \mathbf{x}|\mathcal{K} = \kappa] = \left(\frac{1}{M} \sum_{j=1}^{M} \frac{1}{[\mathbf{y}, \mathbf{x}|\mathcal{K} = \kappa, \theta_j]} \right) \tag{4.14}$$

where θ_j denotes the vector of β and λ_i at the j-th MCMC sample, and M is the total number of MCMC samples. In other words, the posterior distribution for β as given by (4.13) is the weighted average of the posteriors for each κ where the weights are the harmonic means of the likelihoods over the MCMC samples. Assuming a generic parameter sample θ, these likelihoods can be written as

$$[\mathbf{y}, \mathbf{x}|\mathcal{K} = \kappa, \theta] = [\mathbf{y}, \mathbf{x}|\mathcal{K} = \kappa, \beta, \lambda_1, \ldots, \lambda_{n_{\text{basis}}}][\lambda_1, \ldots, \lambda_{n_{\text{basis}}}|\mathcal{K} = \kappa] =$$

$$\prod_{i=1}^{n_{\text{basis}}} \mathcal{N}(y_i^* | \mathbf{1}\{i \leq \kappa\}\beta x_i^*, \lambda_i)\mathcal{N}(\log(\lambda_i) | \log(\hat{\lambda}_i), 1) \tag{4.15}$$

using Equation (4.11) and assuming flat prior on β. Note that here the dependence of λ_i and β on κ is rendered implicit for notational clarity. Regardless, the dependence of λ_i on κ is indirect as they do not explicitly interact in Equation (4.15). The value of κ does directly influence the term $\prod_{i=1}^{n_{\text{grid}}} \mathcal{N}(\mathbf{y}_i^* | \mathbf{1}\{i \leq \kappa\}\beta \mathbf{x}_i^*, \lambda_i)$, which in practice ends up being the predominant factor in driving the distribution over κ. Intuitively, this method will prefer a number of components that maximizes the regression likelihood between $\mathbf{y}_i^*$ and $\mathbf{x}_i^*$.

The primary issue with the BMA approach to selecting κ is that its credible intervals tend to yield sub-nominal coverage rates when evaluated using climate model historical runs as surrogate observations. These results are discussed in detail in Section 4.3.3, but in summary, it appears that the normal likelihood term which drives Equation (4.15) leads to "over-fitting" the data and selecting more than an optimal number of components. This results in suboptimal estimation of β due to the fact that while higher-number components may increase the strength of the fit from the perspective of the likelihood model, if these components correspond to non-meaningful "noise" vectors their inclusion will lead to poor accuracy at recovering the true parameter. At the same time, increasing the number of components increases the model's confidence in its estimation of β, which results in sub-nominal coverage rates. These observations motivate the proposed model for $\mathcal{K}$ in the following section.

4.2.7 Bayesian Model for Component Selection

The BMA procedure for modeling $\mathcal{K}$ uses weights given by Equation (4.15), which prioritizes better fits of regression likelihood term $(y_i^* - \beta x_i^*) \sim N(0, \lambda_i)$. As discussed in Section 4.2.1, accurate estimates of the covariance components are essential to producing reliable estimates and uncertainty intervals. The proposed solution for obtaining a distribution over $\mathcal{K}$ that models the fit of the covariance parameters is through a reparameterization of the data

distributional assumption of $\forall_{i=1}^{\kappa}(\mathbf{y}_i^* - \beta\mathbf{x}_i^*) \sim N(0, \lambda_i)$ to $\sum_{i=1}^{\kappa} \frac{(\mathbf{y}_i^* - \beta\mathbf{x}_i^*)^2}{\lambda_i} \sim \chi^2_\kappa$. This yields the following conditional likelihood model for $\mathcal{K}$:

$$P(\mathcal{K} = \kappa|\mathbf{y}, \mathbf{x}, \theta) \propto [\mathbf{y}, \mathbf{x}|\mathcal{K} = \kappa, \theta] = \chi^2\left(\mathbf{r}_\kappa^T \mathbf{r}_\kappa | df = \kappa - 1\right) \tag{4.16}$$

where the distribution of $\mathbf{r}_\kappa = \Lambda_\kappa^{-1/2} B_\kappa^T(\mathbf{y} - \beta\mathbf{x})$ is obtained from Equation (4.12); note here that the dependence of $\mathbf{r}_\kappa$ on the parameter vector $\theta = (\beta, \lambda_1, \ldots, \lambda_\kappa)$ is rendered implicit. This re-parameterization is inspired by the residual consistency test of Allen and Tett [1999] which yields higher likelihoods when then squared residuals and the component variances are of similar magnitude.

While this reparameterization is motivated by the need to accurately model $\mathcal{K}$, it would be misleading to use this likelihood model for β itself since the mode of (4.16) over β would no longer have a least-squares interpretation. As such, the proposed method of inference is to fit two inter-dependent Bayesian models, one for $\mathcal{K}$ and the other for θ, each of which is conditioned on the value of the other. Under this separation, there becomes two choices for the distribution over $\mathcal{K}$, namely the normal regression parameterization and the χ^2 reparameterization. The former choice yields the following posterior distribution for $\mathcal{K}$ when conditioned on θ:

$$P(\mathcal{K} = \kappa|\mathbf{y}, \mathbf{x}, \theta) \propto \prod_{i=1}^{n_{\text{grid}}} \mathcal{N}(y_i^*|\mathbf{1}\{i \leq \kappa\}\beta x_i^*, \lambda_i). \tag{4.17}$$

The latter choice yields the posterior distribution in Equation (4.16). In the following, the first option of Equation (4.16) will be denoted as $\chi^2(\mathbf{r}_\kappa^T \mathbf{r}_\kappa | df = \kappa) \equiv \ell_{\chi^2}(\kappa, \theta)$ and the second option of Equation (4.17) will be denoted as $\prod_{i=1}^{n_{\text{grid}}} \mathcal{N}(y_i^*|\mathbf{1}\{i \leq \kappa\}x_i^*, \lambda_i) \equiv \ell_{\mathcal{N}}(\kappa, \theta)$. where the dependence on $\mathbf{y}$ and $\mathbf{x}$ is made implicit. These two choices will be generically denoted as elements of the set $\ell \in \{\ell_{\chi^2}, \ell_{\mathcal{N}}\}$. The $\ell_{\mathcal{N}}$ case can be viewed as equivalent to the BMA procedure if a flat prior was to be used for each λ_i.

In summary, the proposed method involves two Bayesian models each conditioned on the other; the first is the regression model for $\theta = (\beta, \lambda_1, \ldots, \lambda_\kappa)$ conditioned on $\mathcal{K} = \kappa$ defined by Equation (4.11) and the second is the model for $\mathcal{K}$ conditioned on θ defined by

either Equation (4.17) in the $\ell_{\mathcal{N}}$ case or by Equation (4.16) in the ℓ_{χ^2} case. The posterior distributions for each of these models can be evaluated simultaneously through the following iterative procedure. As input, this procedure takes the observational field $\mathbf{y}$, the forced field $\mathbf{x}$, a set of basis vector B taken to be either principal components or Laplacian eigenfunctions, empirical variance terms $\hat{\lambda}_1, \ldots, \hat{\lambda}_{n_{\text{basis}}}$ calculated from control vectors, a choice of likelihood function for $\mathcal{K}$ out of the two choices $\ell \in \{\ell_{\mathcal{N}}, \ell_{\chi^2}\}$, a function f which returns a value from a distribution, and a number of MCMC samples M. The method proceeds as follows:

1. Fix a choice of likelihood function ℓ, which generically denotes either $\ell \equiv \ell_{\mathcal{N}}$ or $\ell \equiv \ell_{\chi^2}$. Let $\theta^{(0)} = (\beta^{(0)}, \hat{\lambda}_1, \ldots, \hat{\lambda}_{n_{\text{basis}}})$ denote the "first guess" values for the regression parameters. Using these values, evaluate the distribution $P(\mathcal{K} = k|\mathbf{y}, \mathbf{x}, \theta^{(0)}) \propto \ell(\kappa, \theta^{(0)})$ for each $1 < \kappa \leq n_{\text{basis}}$; note that the restriction $1 < \kappa$ is imposed to avoid degeneracy in the estimation of β.

2. Let $\kappa^{(0)} = (g)P(\mathcal{K}|\mathbf{y}, \mathbf{x}, \theta))$ where g is a function chosen in advanced for providing a value from a distribution. Using MCMC, obtain M samples from the posterior distribution $[\beta|\mathcal{K} = \kappa^{(0)}, \mathbf{y}, \mathbf{x}]$. These samples will be denoted as $\{\beta\}_{j=0}^{M}$.

3. Obtain the point estimates $\beta^{(1)} = g\{\beta\}_{j=0}^{M})$ and similarly $\theta^{(1)}$. Using these values, repeat step 1 to evaluate the distribution $P(\mathcal{K} = k|\mathbf{y}, \mathbf{x}, \theta^{(1)})$ and obtain $\kappa^{(1)} = gP(\mathcal{K}|\mathbf{y}, \mathbf{x}, \theta))$.

4. Iterate steps 2 and 3 until convergence in the posteriors for κ and β.

5. Return posterior samples obtained after convergence.

In the above procedure, each model is fit conditioned on a single value given by the function g. When g gives a random sample from the given distribution, and the posteriors are combined using Equation (4.14), this procedure is equivalent to Bayesian model averaging. While Bayesian model averaging was used in KHS17, such a procedure would not be guaranteed to converge when the ℓ_{χ^2} likelihood function is used since the procedure would be

estimating two inter-dependent posterior distributions simultaneously. This would not be an issue when the $\ell_{\mathcal{N}}$ likelihood is used due to the fact that in this case the compatibility between the likelihood functions for κ and the other parameters means that there is a single posterior target distribution. Nevertheless, in the remainder of the paper g is taken to give the maximum a-posterior (MAP) estimate from a posterior distribution. This renders a procedure which nearly always converges quickly due to the fact that convergence in $gP(\mathcal{K}|\mathbf{y}, \mathbf{x}, \theta))$ implies convergence in the posterior distribution for the other parameters. In the following, the value of $gP(\mathcal{K}|\mathbf{y}, \mathbf{x}, \theta))$ at convergence will be referred to as κ^{post}.

4.3 Statistical Validation using Climate Models

This section will use climate models to evaluate the statistical properties of the proposed innovations, namely the difference between using principal component basis functions versus Laplacian basis functions for parameterizing the covariance matrix and the choice of $\ell_{\mathcal{N}}$ versus ℓ_{χ^2} when modeling $\mathcal{K}$. The combination of principal component basis functions and the $\ell_{\mathcal{N}}$ likelihood model for $\mathcal{K}$ can be seen as analogous to the BMA approach of KHS17, with key differences being the lack of an additive diagonal term, the two-fit procedure for evaluating the posterior as opposed to BMA, and the omission of observational uncertainty. Since these differences complicate the interpretation in comparing the results, the validation study only includes the principal component-$\ell_{\mathcal{N}}$ method and not the BMA approach as described in KHS17. The principal component-ℓ_{χ^2} combination was evaluated to examine the effect of using the ℓ_{χ^2} likelihood when the covariance matrix parameterization is held constant. Finally, the proposed approach using Laplacian basis vectors with the ℓ_{χ^2} likelihood function is evaluated. The combination of Laplacian basis functions with the $\ell_{\mathcal{N}}$ likelihood is omitted here as this method generally performs worse than the others. This is due to the fact that the $\ell_{\mathcal{N}}$ likelihood is more likely to select a larger number of components, which when applied to the Laplacian parameterization leads to the inclusion of components representing

small correlation length scales which are "noisier" from the perspective of inference.

4.3.1 CMIP6 Data

To evaluate these methods, climate model simulation data was used from the 6th version of the Climate Model Inter-comparison product (CMIP6) [Eyring et al., 2016]. To estimate the covariance structure, pre-industrial control (piControl) scenarios were used, and for the forced component historical simulations were used. The former scenarios are run without any greenhouse gas forcing present and as such it can be assumed that they are distributed according to natural variability. The latter scenarios are run with greenhouse gases at historical concentrations with other forcings such as aerosols, ozone, volcanoes, and solar cycles also present. Monthly "tas", or near-surface air-temperature, datasets were used. When necessary datasets were re-gridded to the uniform $5^\circ \times 5^\circ$ latitude-longitude grid over the Earth's surface. Control runs were broken into 25-year segments, and for each segment 25-year trend coefficients were computed at each grid-cell using linear regression. For historical runs regression coefficients were computed over the period 1990-2015.

Denote pre-industrial control and historical climate models as $\mathcal{P}_c$ and $\mathcal{H}_f$ respectively, where $1 \leq c \leq N_{\mathcal{P}}$ and $1 \leq f \leq N_{\mathcal{H}}$. Processing the climate model data yields trend fields $\mathcal{P}_c = \{\mathbf{z}_1, \ldots, \mathbf{z}_{n_{\mathcal{P}_c}}\}$ and $\mathcal{H}_f = \{\mathbf{x}, \ldots, \mathbf{x}_{n_{\mathcal{H}_f}}\}$ for each c and f. Attention has been restricted to model scenarios where $n_{\mathcal{P}_c} > 5$ and $n_{\mathcal{H}_f} > 5$ to ensure that there is sufficient data for testing the inference procedure. Under this restriction $N_{\mathcal{H}} = 7$ and $N_{\mathcal{P}} = 16$ as shown in Table 4.1. Over the 7 historical models, there is a total of $\sum_{f=1}^{N_{\mathcal{H}}} n_{\mathcal{H}_f} = 266$ trend fields in total.

Although individual ensemble members will be used in place of the observations in the validation study described in the next section, ensemble means will be used for estimating the forced pattern. For a fixed $\mathcal{H}_f$, let $\overline{\mathbf{x}} = \frac{1}{n_{\mathcal{H}_f}} \sum_i x_i$ denote the ensemble mean and let $\overline{\mathbf{x}}_{-k} = \frac{1}{n_{\mathcal{H}_f}-1} \sum_{i \neq k} \mathbf{x}_i$ denote the mean over the ensemble with the k-th member excluded. Figure 4.5 displays the historical trend fields for each of the seven historical simulations.

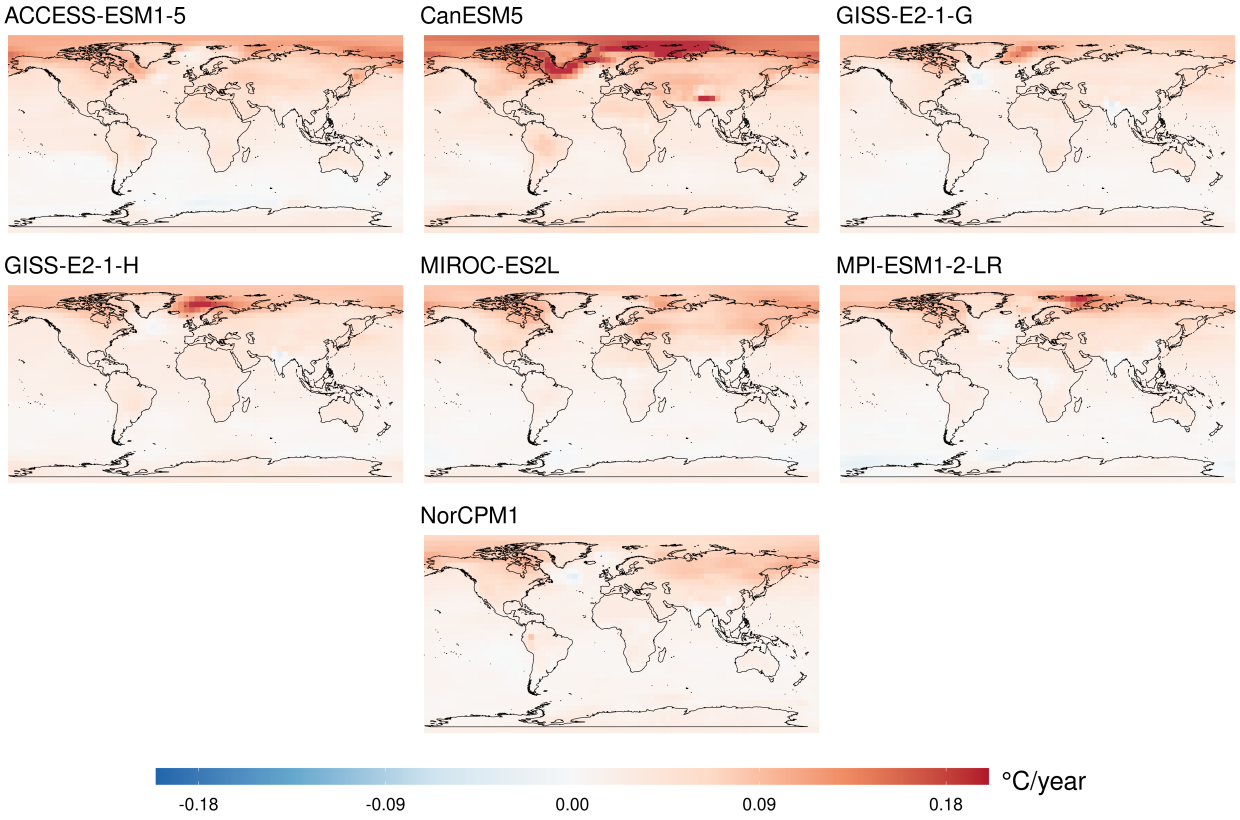


Figure 4.5: Mean forced signal for each of the 7 historical climate model ensembles.

The variation between these estimated forcing patterns implies an additional source of uncertainty arising from error in the climate simulations, which is in addition to the uncertainty induced by estimating the forced signal from limited runs. As historical ensemble means are used, the latter source of uncertainty is not accounted for here, although it would be straightforward to incorporate this portion into the hierarchical Bayesian model in a way similar to KHS17. While simulation error is also not accounted for in the proposed approach, it will be evaluated in the results of this section through intervals over quantities obtained from applying the statistical frameworks to the different historical models. This is similar to "simulation error" implied by heterogeneity in the implied covariance structures over $\mathcal{P}_c$ and visualized in Figures 4.2 and 4.4, which will also be evaluated in the following study.

4.3.2 Study Design

For a fixed historical climate model $\mathcal{H}_f$, for any $\mathbf{x}_k \in \mathcal{H}_f$ the value of the regression parameter β in the Equation $\mathbf{x}_k = \beta \mathbf{x}_{\mathcal{H}_f} + \epsilon$ is known to be equal to one. This "true value" setting can be used to evaluate the statistical properties of the Bayesian optimal fingerprinting procedure using historical vectors as a surrogate for the observations. To ensure that the true value of β is one, the climate model generating the surrogate observations will be kept the same as the model used for estimating the forced signal. However, for each historical model, the climate model generating the pre-industrial control vectors will be alternated over the 16 choices in Table 4.1. This alternating scheme is done in order to investigate the case where the covariance matrix implied by the control vectors differs from the "true" covariance generating the observations. The hypothetical "true" covariance matrix describing the real-world observations can be thought of as being at least as different from any particular climate model's covariance structure as the covariance structure of any two climate models are from each other. As such, varying the choices of historical and pre-industrial climate models can yield insight into the statistical properties of applying the method with a particular pre-industrial control model to real-world observations.

For the study, statistical properties will be evaluated in terms of coverage rates, root-mean-squared errors (RMSE), and continuous ranked probability scores (CRPS), where the CRPS metric evaluates both the accuracy and precision of the predictive distribution [Hersbach, 2000]. The validation procedure described below takes as input a choice of either principal component or Laplacian basis vectors, a choice of likelihood function for $\mathcal{K}$ out of the choices ℓ_{χ^2} or $\ell_{\mathcal{N}}$, and choices for g and M as required by the two-fit procedure. With these the study proceeds as follows:

1. Choose set of pre-industrial control vectors $\mathcal{P}_c$ for $1 \leq c \leq N_{\mathcal{P}}$.

2. Choose set of historical vectors $\mathcal{H}_f$ for $1 \leq f \leq N_{\mathcal{H}}$

3. For each $1 \leq f \leq n_{\mathcal{H}_f}$, let $\mathbf{y} \equiv \mathbf{x}_k$ represent the surrogate bservations and let $\mathbf{x} \equiv \overline{\mathbf{x}}_{-k}$, where $\overline{\mathbf{x}}_{-k}$ is the average of the $n_{\mathcal{H}_f} - 1$ ensemble members excluding the k-th. The vector $\overline{\mathbf{x}}_{-k}$ represents the estimate of the true forced signal $\mathbf{x}$.

4. Given this (c, f, k) pair, use the surrogate observations, estimated forced pattern, and pre-industrial control runs to fit the posterior distribution $[\beta|\mathbf{x}, \mathbf{y}, \kappa^{\text{post}}]$ according to the two-fit procedure described in Section 4.2.7. Recall that κ^{post} is the MAP value of the posterior distribution for $\mathcal{K}$ at convergence.

5. From the resulting posterior distribution of β, calculate $CI_{c,f,k} = \mathbf{1}\{1 \in [q_{\beta,0.5}, q_{\beta,.95}]\}$ where $q_{\beta,0.5}$ and $q_{\beta,.95}$ denote quantiles. This will be equal to 1 if the 90% posterior credible interval for β includes the true value of $\beta = 1$. Additionally, calculate the posterior mean of β, which will be denoted as $\beta_{c,f,k}$, and the CRPS value for the posterior distribution of β against the true value of $\beta = 1$.

6. After fitting the model for each $1 \leq k \leq n_{\mathcal{H}_f}$, calculate the coverage rate $\overline{CI}_{c,f} = \frac{1}{n_{\mathcal{H}_f}} \sum_k CI_{c,f,k}$. Also calculate the RMSE scores over posterior means $\text{RMSE}_{ij} = \sqrt{\frac{1}{n_{\mathcal{H}_f}} \sum_k (\beta_{c,f,k} - 1)^2}$.

7. After obtaining $CI_{c,f}$ for each historical model $\mathcal{H}_f$, obtain the distribution of coverage rates $\mathcal{CR}_c = \{CI_{c,f}\}_j$ as well as the analogous distributions over CRPS and RMSE scores.

The above steps are repeated until $\mathcal{CR}_c$, as well as distributions over RMSE and CRPS values, are obtained for each $1 \leq c \leq N_{\mathcal{P}}$.

4.3.3 Validation Study Results

The above-described validation study was run with each MCMC fit having $M = 2{,}000$ samples and burn-in period of 1,000 iterations. The function g was taken to yield maximum a-posteriori (MAP) point-estimates for the reasons described at the end of Section 4.2.7, although the same reasoning would also justify the use of posterior medians or means. The study was run for three sets of modeling choices; principal components and $\ell_{\mathcal{N}}$, principal components and ℓ_{χ^2}, and Laplacian components and ℓ_{χ^2}. In reported results, the principal component basis function approach will be referred to as empirical orthogonal function (EOF) approaches, EOF being a common name for principal components in the optimal fingerprinting literature. Overall, for each of the three sets of modeling choices, this validation study required $N_{\mathcal{P}} * \sum_j n_{\mathcal{H}_f} = 16 * 266 = 4256$ two-step model fits over the (c, f, k) tuples.

Figure 4.6 displays the distribution of κ^{post} over the (f, k) tuples for each pre-industrial climate model $1 \leq c \leq N_{\mathcal{P}}$. Distributions are visualized through means and 5-th and 95-th intervals calculated over the historical ensemble members plotted on the $\log_2$ scale. The models have been ordered on the x-axis by increasing values of $n_{\mathcal{P}_c}$. For the two EOF basis function fits, the number of pre-industrial control runs is shown with black lines, which form a cap on the maximum value of κ^{post}. The maximum potential number of Laplacian basis functions is $n_{\text{grid}} = 2592$, which is beyond the limit of the y-axis.

It can be seen that the EOF-$\ell_{\mathcal{N}}$ version tends to put weight on a larger number of components, and there is a clear trend in increasing κ^{post} distributions as $n_{\mathcal{P}_c}$ increases.

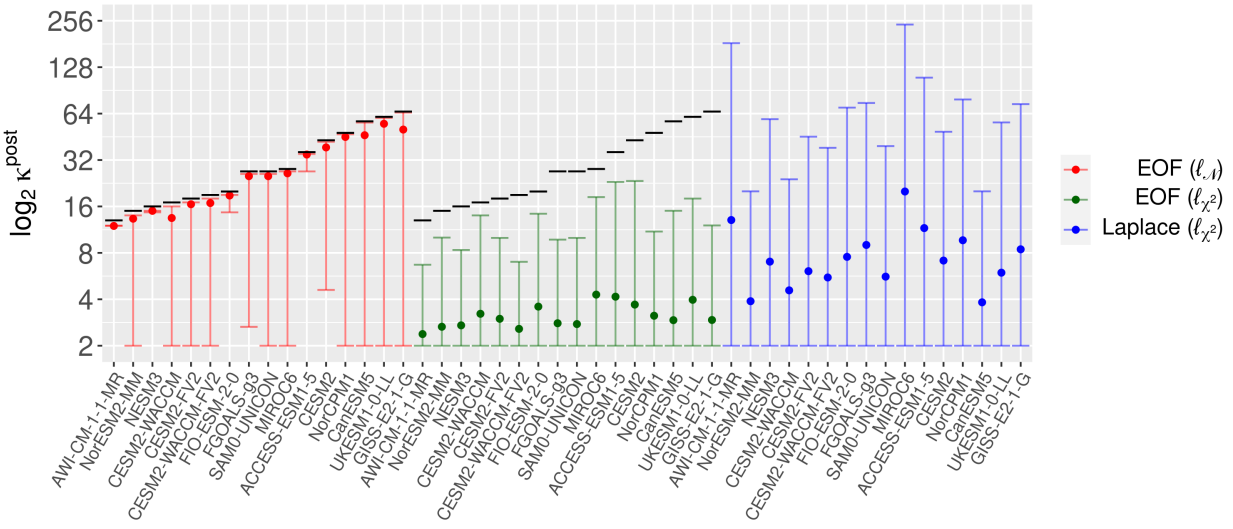


Figure 4.6: Distribution of κ^{post} plotted on the $\log_2$-scale (y-axis) over each pre-industrial control model (x-axis) for each of the three statistical frameworks described in Section 4.3.2. Means and 90%-intervals are shown over the set of all historical ensemble members across the different pre-industrial models. For the EOF basis function approaches, κ^{post} is capped at $n_{\mathcal{P}_c}$, or the number of pre-industrial control vectors, which are shown as black lines. Pre-industrial models are ordered according to increasing number of control vectors $n_{\mathcal{P}_c}$.

This increasing trend is not the case with the ℓ_{χ^2} versions. It can be seen that in the EOF-ℓ_{χ^2} version, the number of basis functions is generally quite small, with median values often less than eight and always less than the median value under the $\ell_{\mathcal{N}}$ selection method. For each pre-industrial model, the lower-bound on κ^{post} reaches 2, which is the smallest number of basis functions allowed in the implementation. The Laplace-ℓ_{χ^2} method generally selects more components on average than the EOF-ℓ_{χ^2} method and either more or less than those with the EOF-$\ell_{\mathcal{N}}$ method, although it should be noted that since these use different sets of basis functions the number chosen is not directly comparable. As an example, when $\kappa^{\text{post}} = 2$, the EOF method will project on components that differ according to the pre-industrial control model used, the first of which can be seen in Figure 4.2. When $\kappa^{\text{post}} = 2$ with Laplacian basis functions the method will project onto the global mean component and onto the north-south variation component seen in Figure 4.3. In addition, since by construction the EOF approach will maximize the amount of variance explained by each successive component, any number of EOFs will explain an amount of variance that would require a larger number of Laplacians. As such, it is not surprising that a smaller number of EOFs are selected than Laplacians when the adequacy of the covariance estimation is prioritized with the ℓ_{χ^2} parameterization.

Figure 4.7 shows the distributions of $\mathcal{CR}_c$ for each $1 \leq c \leq N_{\mathcal{P}}$, recalling that $\mathcal{CR}_c$ is the collection of coverage rates for the 90% Bayesian credible intervals when the "true" β is known to be one. These are displayed using medians and interquartile ranges over the distribution implied by the historical models, and a horizontal line shows the desired coverage rate of 90%. As in Figure 4.6, the control models on the x-axis are ordered by increasing number of available control vectors.

For the EOF-$\ell_{\mathcal{N}}$ method, it can be seen that coverage rates are almost always less than 90% and that under-coverage continues to worsen as $n_{\mathcal{P}_c}$ increases. This property of the EOF-$\ell_{\mathcal{N}}$ method, where the coverage rates decline as the number pre-industrial control models increase, is undesirable as an ideal method would increase in reliability as more data

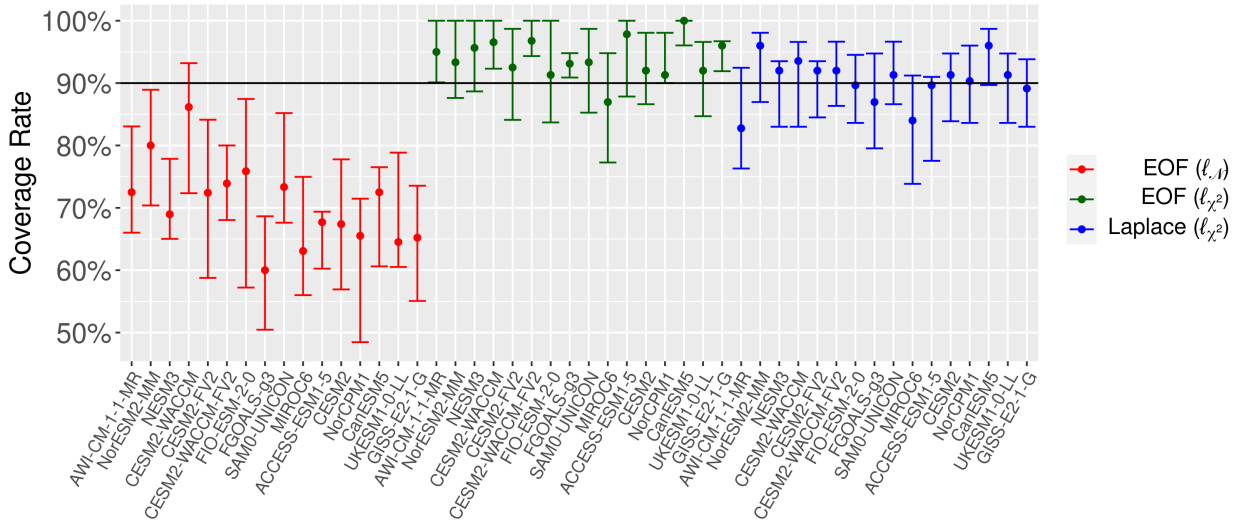


Figure 4.7: Distribution of $\mathcal{CR}_c$, or the percentage of 90% credible intervals that contain the true value of $\beta = 1$, for each pre-industrial control model. Coverage rates were calculated for each historical model separately, and means and 90%-intervals are shown over the set of all historical ensemble members. Pre-industrial models on the x axis are ordered according to increasing number of control vectors.

becomes available. From Figure 4.6 it can be seen that as the number of pre-industrial control vectors increases, this selection approach is more likely to estimate a larger number of components. This leads to inaccurate coverage intervals as the $\ell_{\mathcal{N}}$ likelihood function "over-fits" the number of components. As discussed in Section 4.2 this leads to the selection of non-meaningful components that, while improving the fit from the perspective of the likelihood function, leads to over-confidence and sub-optimal accuracy as the number of components gets larger.

Results for the EOF-ℓ_{χ^2} likelihood often achieve accurate coverage but also appear to have a slight pattern of over-coverage. However, the Laplace-ℓ_{χ^2} method achieves well-calibrated coverage rates that do not depend on $n_{\mathcal{P}_c}$, with all of the sixteen intervals including 90%. Recall that the ℓ_{χ^2} distributional assumption for $\mathcal{K}$ is intended to prioritize accurate modeling of the covariance components. This choice produces notably better-calibrated coverage rates, with the Laplacian basis parameterization providing a smaller additional improvement over the principal component approach.

Figure 4.8 displays the root mean squared error (RMSE) scores calculated for each historical climate model $\mathcal{H}_f$ using the posterior mean values of β from each ensemble member k. Medians and interquartile ranges over $\mathcal{H}_f$ are displayed for each pre-industrial climate model $\mathcal{P}_c$ on the x-axis. It can be seen that, in addition to having well-calibrated coverage rates, the Laplace-ℓ_{χ^2} method generally has the lowest RMSE scores out of the three model selection versions, indicating that this method is able to identify a covariance structure that not only produces accurate recovery of β but also produces accurate uncertainty values. The EOF-ℓ_{χ^2} has the worst RMSE scores on average and the largest spread both within and between pre-industrial control models. This is unsurprising as this method will generally select a small number of components that are not optimal for estimating β due; this will be discussed in further depth below. The heterogeneity of the first principal components of the covariance structures for each pre-industrial model explains the larger degree of variability in the results for the two EOF methods. The EOF-$\ell_{\mathcal{N}}$ method has a somewhat lower spread

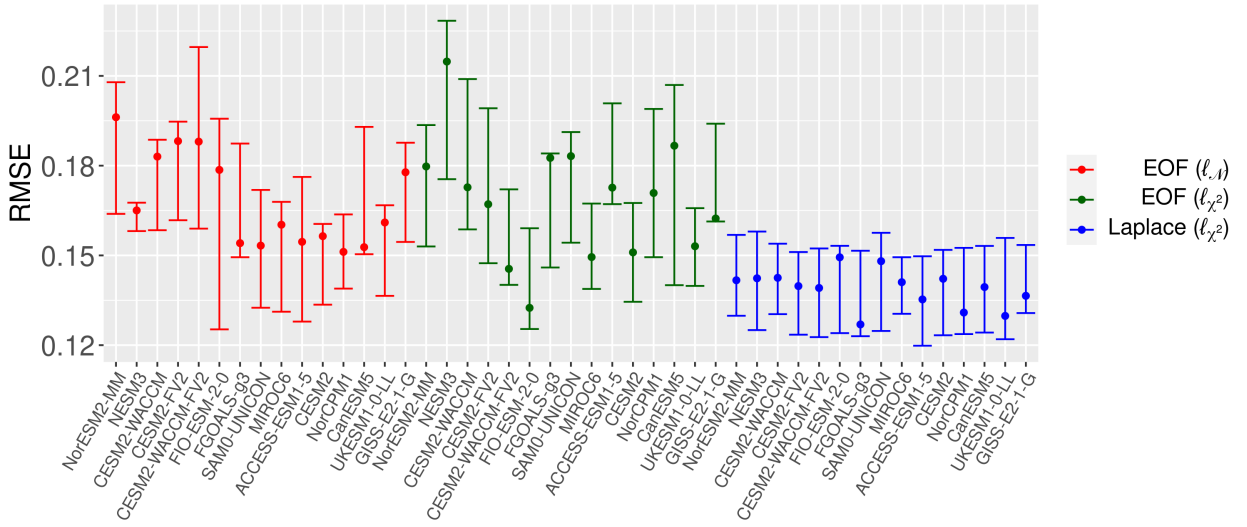


Figure 4.8: Distribution of RMSE scores calculated using the posterior means of β for each ensemble member. Scores are calculated against a "true" value of $\beta = 1$. Medians and interquartile ranges are shown over the historical climate models for each pre-industrial control model on the x-axis.

and improved scores over the EOF-ℓ_{χ^2} model. This lower spread is due to the larger number of components estimated, so the estimated covariance structures are allowed to have a closer match between the historical model-derived surrogate observations and the control. Both of the EOF methods also exhibit improving scores as $n_{\mathcal{P}_c}$ increases, and while it is not surprising that increasing the amount of data results in improved performance for these methods, this feature is notably absent in the Laplace results, which exhibit low RMSE scores across the pre-industrial control models. The Laplacian-ℓ_{χ^2} also has the lowest spread both within and between pre-industrial control models. This is due to the fact that the use of Laplacian basis functions removes both the variability induced by the estimation of principal components from pre-industrial simulation data and the variability induced by the "mismatch" between estimated principal components and the unobserved covariance structures implied by the surrogate observations. CRPS values are displayed in Figure 4.9, which shows the same general patterns.

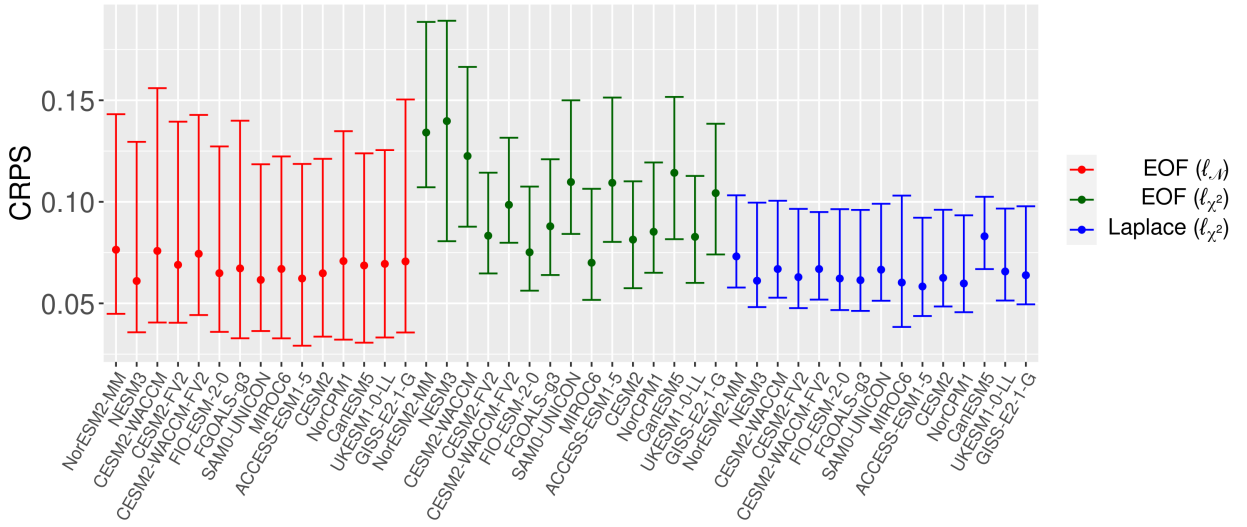


Figure 4.9: Distribution of CRPS values calculated using the posterior means of β for each ensemble member. Scores are calculated against a "true" value of $\beta = 1$. Medians and interquartile ranges are shown over the historical climate models for each pre-industrial control model on the x-axis.

The "mismatch" phenomenon, first described in Section 4.2.3, can explain the difference in performance between the EOF-ℓ_{χ^2} and the Laplace-ℓ_{χ^2} results. Conceptually the observations can be thought of as being generated from a true covariance matrix that is unknown to the observer. Furthermore, it can be assumed that this theoretical climate model has a variability structure that may be at least as different from any particular climate model as the climate models are from each other. While it has been assumed here that the forced and unforced variability structures are the same, in application and in the validation study this mismatch will occur whenever the observations are from a different real or theoretical covariance structure than the pre-industrial control simulation. This will always occur for the observations due to simulation error and is represented in the validation study through all (c, f) pairs where $\mathcal{P}_c$ and $\mathcal{H}_f$ represent different models. The heterogeneity of the implied covariance structures between the climate models can be visualized for the EOF approach in Figure 4.2 and for the Laplace approach in Figure 4.4. This mismatch entails over-estimation

of the variance for lower-number EOFs, since the most significant patterns represented by lower-number EOFs will correspond to less significant and less variable patterns in the observations. This leads to over-coverage when a small number of EOF components are used, as is often the case in the EOF-ℓ_{χ^2} method (Figure 4.7). Another way to look at this is that for any two climate models, the first or highest variance EOF of the first model will correspond to a higher-number and lower variance pattern or combination of patterns in the second model. This explains why over-coverage is only present in the lower-$n_{\mathcal{P}_c}$ models in the EOF-$\ell_{\mathcal{N}}$ results, since selecting higher numbers of components allows for patterns to be matched between different EOF structures. Conversely, this mismatch phenomenon forces the EOF-$\ell_{\mathcal{N}}$ method to over-select the number of components, leading to under-coverage as $n_{\mathcal{P}_c}$ grows.

The use of Laplacian basis functions, on the other hand, does not suffer from mismatching in the structure of the components since they are fixed across historical and pre-industrial models. While under this basis there is still heterogeneity in the component variances as estimated using different climate models, the estimation and uncertainty of these parameters are taken into account in the Bayesian regression framework. This generality and ability to propagate uncertainty in the variance weights yields better-calibrated coverage rates than the two EOF approaches as well as lower RMSE and CRPS values.

A caveat to the use of the Laplacian basis parameterization is that the displayed results only use the ℓ_{χ^2} likelihood for the distribution over $\mathcal{K}$. While not reported here, it appears that when normal likelihood-based component selection methods are used with the Laplacian basis parameterization, the method tends to over-fit the number of components, resulting in under-coverage and worse RMSE scores than the other methods. This issue is not unique to the Laplacian basis parameterization but is due to the larger number of possible components allowed; as can be seen from Figure 4.7 when the $\ell_{\mathcal{N}}$ method is used with the EOF parameterization, under-coverage becomes more severe as the number of possible components increases. When the effective number of components is increased by adding a diagonal term

to the covariance matrix, coverage appears to decrease dramatically. The proposed two-step ℓ_{χ^2} approach provides a solution by modeling the number of components with a likelihood function that prioritizes covariance fit. However, the separation of the model for $\mathcal{K}$ from the regression model is somewhat of an ad-hoc approach that does not have the theoretical justification of an integrated Bayesian model. Practically the separation of the models has the undesirable property that it is not able to propagate uncertainty in the number of components to the final result. Ideally, this two-fit approach will serve as a foundation for an integrated approach that shares the desirable properties of the ℓ_{χ^2} model for $\mathcal{K}$, however this extension is left for future work.

In Figures 4.8 and 4.9, the results from the pre-industrial climate model AWI-CM-1-1-MR have been omitted since its values in several instances exceeded a reasonable range for the y-axis. The results for this model are reported separately in Table 4.2. For this model the Laplace-ℓ_{χ^2} method has the lowest median RMSE and CRPS values as well as the lowest spread in the values, agreeing with the findings of Figure 4.8.

Model	Method	Metric	$q_{.05}$	$q_{.50}$	$q_{.95}$
AWI-CM-1-1-MR	EOF-$\ell_{\mathcal{N}}$	RMSE	0.15	0.16	0.17
AWI-CM-1-1-MR	EOF-ℓ_{χ^2}	RMSE	0.45	0.59	1.05
AWI-CM-1-1-MR	Laplace-ℓ_{χ^2}	RMSE	0.15	0.15	0.17
AWI-CM-1-1-MR	EOF-$\ell_{\mathcal{N}}$	CRPS	0.04	0.08	0.13
AWI-CM-1-1-MR	EOF-ℓ_{χ^2}	CRPS	0.24	0.33	0.61
AWI-CM-1-1-MR	Laplace-ℓ_{χ^2}	CRPS	0.05	0.06	0.12

Table 4.2: Summary of validation study results from the AWI-CM-1-1-MR pre-industrial runs, which were omitted from the figures due to space constraints.

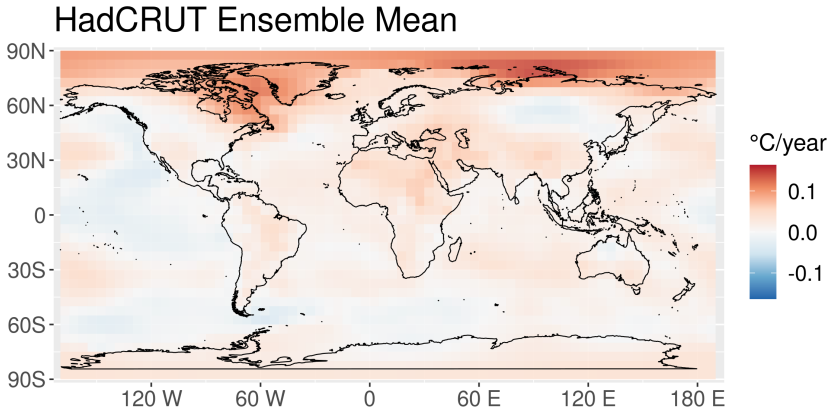


Figure 4.10: HadCRUT observational ensemble mean of near-surface air temperature trends over 1990-2015.

4.4 Application to HadCrut Observations

After evaluating the performance of these three methodological choices within a "known truth" context provided by climate model simulations, these frameworks can be applied to real observations to evaluate the substantive effect of the methodological choices. For the observational data, let $\mathbf{y}$ be the mean 1990-2015 near-surface air temperature trend field of the 200-member HadCRUT global temperature ensemble [Osborn et al., 2021, Morice et al., 2021]. This field is displayed in Figure 4.10. It should be noted that the application presented here does not consider the uncertainties associated with the observational ensemble in order to keep the methods unchanged from the validation study where observational uncertainties were not simulated.

To quantify the effects of heterogeneity in the choice of climate models, the two-step procedure was applied to the observations with each of the $16 * 7 = 112$ pairs of historical and pre-industrial models. Posterior means β_{post} are displayed in Figure 4.11. All of the results tend to have a median values notably less than one. This is somewhat surprising, but the consistency of this result over climate model pairs and methods indicates that this

feature is inherent in the data. It can be seen that the Laplace-ℓ_{χ^2} method exhibits smaller inter-model variability than the other two approaches. An interesting feature is that the median estimates for the EOF-$\ell_{\mathcal{N}}$ get larger as $n_{\mathcal{P}_c}$ increases, and appear to get closer to the fairly consistent median values of the Laplacian results. This is a testament to the fact that while the accuracy of the EOF-$\ell_{\mathcal{N}}$ methods increases with $n_{\mathcal{P}_c}$, the Laplacian method exhibits consistently accurate results regardless of the number of control vectors.

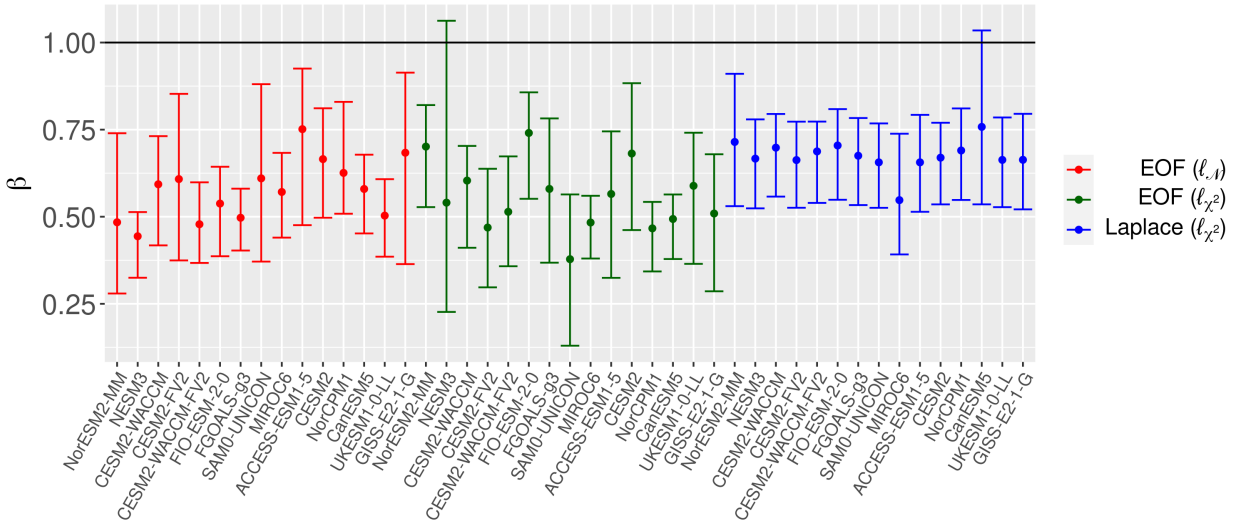


Figure 4.11: Posterior means of β from the application of the Bayesian optimal fingerprinting methods to HadCRUT observational data. Medians and 90% intervals over the historical model runs are displayed. The black horizontal line is at $\beta = 1$.

With β_{post} and posterior standard deviations $\sigma_{\beta;\text{post}}$ the detection problem can be viewed in terms of the ratio $\frac{\beta_{\text{postmean}}}{\sigma_{\beta;\text{post}}}$, or the distance in posterior standard deviations between the posterior mean of β and zero. Greater values of this statistic indicate further evidence for detection. In the hypothesis testing framework, the null hypothesis $H_0 : \beta = 0$ would be rejected at significance level α if $\frac{\beta}{\sigma_\beta} > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$ quantile of the normal distribution. While the displayed results come from a Bayesian framework, similar logic can be employed to make binary detection and attribution conclusions.

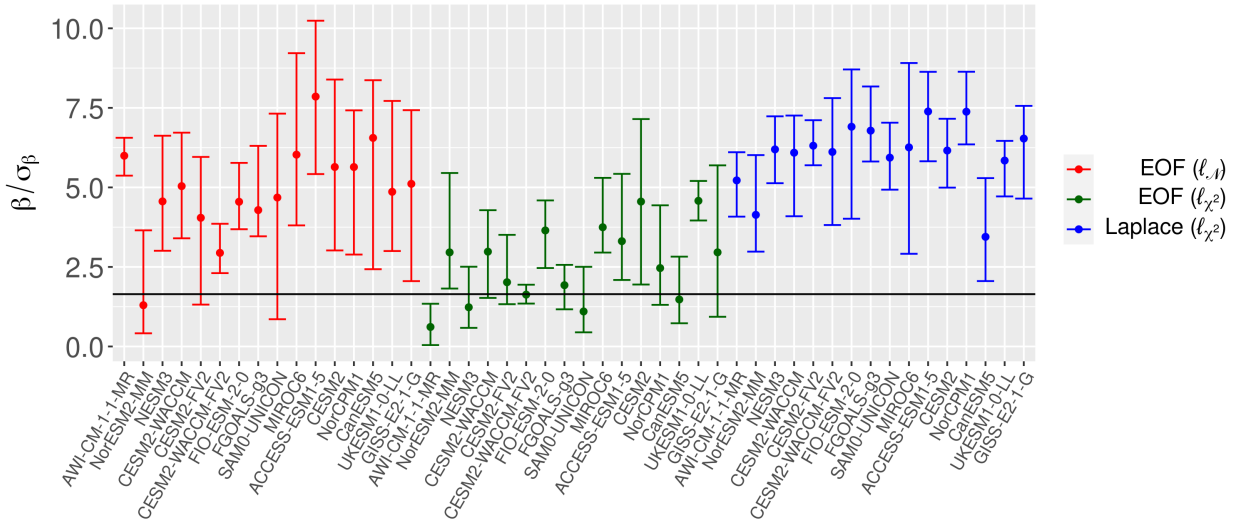


Figure 4.12: Evidence for "detection" of the historical signal in the HadCRUT observational data. The y-axis is presented in terms of number of standard deviations above zero, or $\frac{\beta_{\text{postmean}}}{\sigma_{\beta;\text{post}}}$. Medians and 90% spreads over the historical models are displayed. The black line is at $z_{1-0.05} = 1.64$; values above this line would be considered "detected" at the 5% threshold.

The results for detection are displayed in Figure 4.12. Here, medians and 90% intervals are displayed over the set of historical models for each of the pre-industrial control models. The horizontal line at $z_{1-0.05} = 1.64$ represents the threshold where the probability of β being less than or equal to zero falls below 5%. It can be seen that the EOF-ℓ_{χ^2} method yields substantially different results than the other two methods in that it frequently fails to find strong evidence for detection. This is due to the fact that this method over-estimates the variance, as evidenced by the validation study of Section 4.3. The EOF-$\ell_{\mathcal{N}}$ and Laplace-ℓ_{χ^2} methods both conclude detection over most model pairs, however the Laplacian method exhibits notably less variability between historical and pre-industrial climate models. The higher level of inter-model variability in the EOF approaches is not surprising given the dependence of the EOF representations on the specific control vectors used, and further highlights this advantage of the Laplacian method.

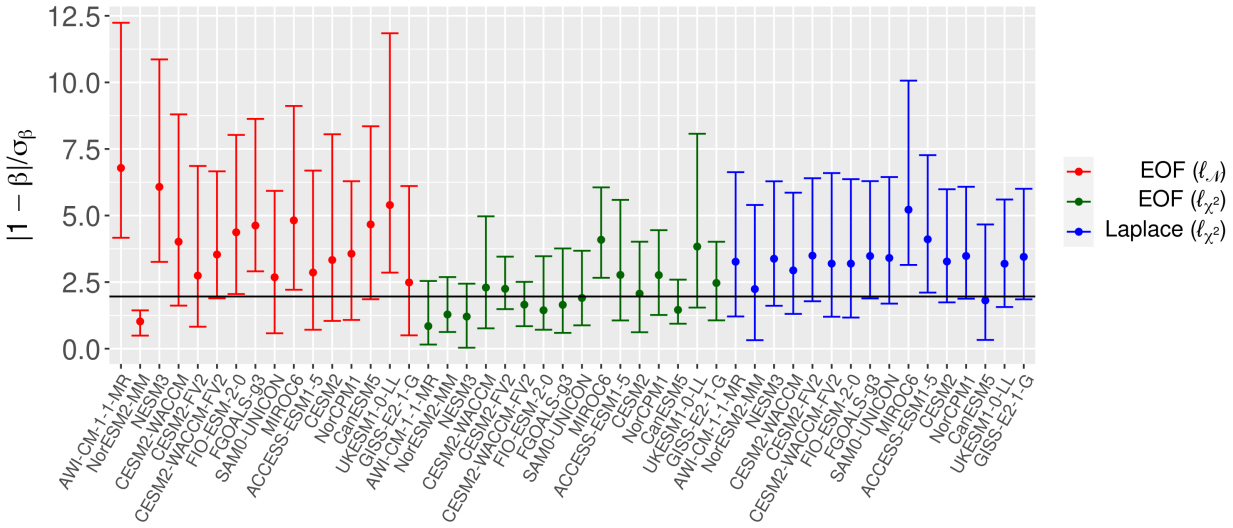


Figure 4.13: Evidence for the "attribution" of trends in the HadCRUT observational data to the historical signal. The y-axis is presented in terms of standard deviations away from one, or $\frac{|1-\beta_{\text{postmean}}|}{\sigma_{\beta;\text{post}}}$. Medians and 90% spreads over the historical model results are displayed. The black line is at $z_{0.05/2} = 1.96$; values above this line would be considered "attributed" at the 5% threshold.

For attribution, the statistic $\frac{|1-\beta_{\text{post mean}}|}{\sigma_{\beta;\text{post}}}$ indicates the distance in posterior standard deviations between the posterior mean and one. This can be used analogously to the frequentist attribution test, where $\frac{|1-\beta|}{\sigma_\beta}$ would be the test statistic for rejecting or failing to reject $H_0 : \beta = 1$ with a two-sided test. The results are shown in Figure 4.13, where it can be seen that while attribution rates are similar for EOF-$\ell_{\mathcal{N}}$ and Laplace-ℓ_{χ^2} methods, the inter-model spread is lower for the Laplace-$\ell\chi^2$ approach. The EOF-ℓ_{χ^2} approach rarely makes attribution conclusions, which is due to the over-estimated variances obtained with this method.

A numerical summary of the detection and attribution results are displayed in Table 4.3. Here, "# Detected" refers to the percentage of the 112 historical-control combinations whose posterior distribution gives less than a 5% probability of β being less than or equal to zero. This can be seen as analogous to rejecting the null hypothesis of $\beta = 0$ at the 5% level within a frequentist framework. It can be seen here that the Laplace-ℓ_{χ^2} method makes a detection conclusion for every model pair, a testament to this method's consistency and accuracy. The third column shows the percentage of "attribution" conclusions, defined as when the 95% credible interval of the posterior distribution contains $\beta = 1$. The Laplace-ℓ_{χ^2} method also has the highest percentage of positive attributions, however none of the attribution rates are particularly high due to the fact that β is generally estimated to be less than one (see Figure 4.11). The third column shows the average value of the detection statistic plotted in Figure 4.12. The Laplace methods has the highest average confidence in a positive trend, with on average the posterior mean being 6.3 standard deviations greater than zero. This test statistic is slightly higher than that of the EOF-$\ell_{\mathcal{N}}$ method, and unsurprisingly the EOF-ℓ_{χ^2} method has the lowest value.

Method	# Detected	# Attributed	Mean $\frac{\overline{\beta}_{\text{post}}}{\sigma_{\beta;\text{post}}}$
EOF ($\ell_{\mathcal{N}}$)	100 (89.3%)	19 (17%)	5.16
EOF (ℓ_{χ^2})	70 (62.5%)	33 (29.5%)	2.13
Laplace (ℓ_{χ^2})	112 (100%)	34 (30.4%)	6.30

Table 4.3: Summary of results applied to HadCRUT observational data over each of the 112 control-historical climate model combinations. "Detection" and "attribution" results were determined using a 5% posterior probability cutoff.

4.5 Conclusion

The use of Laplacian basis functions and the novel approach for modeling the number of components presents a further step forward on the path towards more reliable detection and attribution results. The proposed approach has been motivated by two drawbacks in traditional optimal fingerprinting methodology. The first is the use of principal components or EOFs to parameterize the covariance matrix. The fact that the principal components vary noticeably between different climate models means that their use introduces substantial variability into the optimal fingerprinting procedure which is not taken into account in inference. The proposed solution is to use Laplacian basis functions to provide a flexible, spatially-coherent parameterization of the covariance matrix whose structure does not need to be estimated from climate simulation data. The second issue is the selection of the number of components, which substantially affects detection and attribution results when using either the principal component or Laplacian covariance parameterizations. The validation study demonstrates that selection approaches based on the normal likelihood parametrization for the number of components produce results that are over-confident when the potential number of components is high. The proposed solution is to model the number of components with a separate Bayesian model inspired by the χ^2 residual consistency test. This yields inter-dependent models which can be simultaneously fit in an iterative fashion. When

the normal likelihood is used for both models, the two-fit procedure can be viewed as equivalent to an integrated Bayesian approach. The "true data" climate model study has shown that the Laplacian parameterization using the χ^2 model for the number of components has better-calibrated coverage rates and lower error scores than the principal component parameterization. Finally, these techniques have been demonstrated on observational data, where the Laplace-χ^2 method detects the anthropogenic signal in all but one of the 112 climate model combinations with an average confidence of 5.16 standard deviations.

A notable drawback to this approach is the ad-hoc nature of the two-fit method. The two separate but inter-dependent models are each independently able to quantify uncertainty but do not able to propagate uncertainty through each other. This could be solved through an integrated Bayesian framework for both model selection and parameter estimation. Such an approach is taken in the Bayesian framework of KHS17, however in that case the posterior distribution for the number of principal components is based on the normal likelihood model, which has been shown in Section 4.3.3 to give sub-optimal coverage rates. Future work will investigate the potential for a Bayesian framework that can combine the two models proposed here in an integrated fashion.

Another limitation to the proposed approach is that the model selection technique only allows for consecutive sets of components to be included. Such an approach is sensible in the principal component basis parameterization, where successive components have lower empirical variances. In the Laplacian case, however, the relationship of variance to component number is not monotonic as can be seen in Figure 4.4. Intuitively, when looking at the first nine components in Figure 4.3 it could, for example, be sensible to include the first, second, and fifth components (the mean, north-south, and pole-equatorial patterns respectively) while excluding the more longitudinally correlated third and fourth patterns. One potential future approach to address this would be to have a model selection parameter for each component, although this would come with the natural difficulty in that the joint posterior distributions for such parameters would be high-dimensional when the number of

potential components is large. Another approach could be to re-order the Laplacians prior to fitting the model, although this would have the undesirable effect of introducing a new source of variability that would not be accounted for in the framework.

This chapter has focused on quantifying uncertainty in the covariance matrix through the form and number of basis functions. To this end other important sources of uncertainty that are accounted for in other optimal fingerprinting approaches have been ignored, including uncertainty in estimating the forced component and observational uncertainty. Both of these sources of estimation uncertainty can be incorporated into the Bayesian model presented here in a straightforward fashion. A more challenging source of uncertainty is climate model error. In the validation study and application, the methods were applied to only a singular pair of historical and pre-industrial control climate models at a time, and the results show a considerable amount of variability across climate model choice. The observations can be viewed as being generated by a covariance structure which is unknown to us but is assumed to be a member of the distribution of covariance structures implied by the spread of inter-model variability. Under this perspective a Bayesian model could be developed which incorporates data from multiple climate models and is able to simultaneously quantify within-model estimation uncertainty and between-model structural uncertainty. The Laplacian covariance matrix parameterization provides a starting point for such a model, as the fact that their form is constant across climate models means that the covariance structures can be pooled within the context of a hierarchical Bayesian framework. This extension has the potential to considerably increase the robustness of the optimal fingerprinting methodology presented here and is planned to be pursued in future work.

CHAPTER 5

Discussion and Directions for Future Work

The three statistical themes introduced in Chapter 1 all play a crucial role in the methods developed over the previous three chapters. The first theme, regarding the use of spatially-coherent covariance structures to model climate phenomena, was represented in the cylindrical kernel-convolution Gaussian process used in the ocean heat content problem and in the Laplacian basis parameterization used in the detection and attribution problem. It is worth noting that a major reason why spatially-coherent covariance structures are infrequently used in the climate literature is the computational burden often presented by traditional methods. In the ocean heat content problem, the computational burden of the Gaussian process is particularly prescient, as computing the Cholesky decomposition of a covariance matrix on the over one million observations of the full dataset would require on the order of 1×10^{18} operations. This problem is alleviated through the use of a computationally efficient Vecchia process, which was shown in Section 2.5 and Section 3.5 to provide an adequate approximation for the full Gaussian process model. In the detection and attribution problem, the computational difficulty presented by spatial covariance models is alleviated by using a Laplacian basis structure rather than a Gaussian process structure. While covariance matrices parameterized using Laplacian basis functions can represent patterns of spatial correlation similar to Gaussian processes, the Laplacian basis functions are not dependent on an additional range parameter and as such can be computed in advance. Gaussian processes, even when using the Vecchia approximation, must be re-computed for every configuration of the parameters, although recent work in non-parametric Vecchia processes for climate covariance matrices could potentially render these methods suitable for the detection and

attribution problem in future work [Kidd and Katzfuss, 2021].

The second theme, regarding the use of hierarchical Bayesian models to propagate complex sources of uncertainty to the final results, is used in both problems to represent the uncertainty in the estimation of the covariance structure. In the ocean heat content problem, the Bayesian hierarchical model developed in Chapter 2 represents the covariance parameters as Gaussian processes themselves to allow for the propagation of their uncertainty to the final results. In the detection and attribution problem, the coefficients of the Laplacian basis parameterization were incorporated into a Bayesian hierarchical model in order to propagate the uncertainty in their estimation to the final results. Furthermore, an additional layer was added to the hierarchy involving a chi-squared reparameterization of the likelihood function to obtain a more reliable estimation of the number of Laplacian basis functions to use.

The third theme, regarding the statistical evaluation of the posterior distributions, is exhibited in both the ocean heat content and the detection and attribution problems, where both proposed frameworks have been shown to achieve better statistical properties than previously used approaches. For the ocean heat content model, two different types of cross-validation were used to evaluate how well the posterior distributions represent the process at unobserved locations. In addition to the commonly used metrics of root-mean-squared error and mean absolute error, continuous-ranked probability scores were used in order to evaluate both the mean and spread of the posterior distributions. For the detection and attribution problem, the framework was evaluated using climate model simulations to evaluate the method under the situation where the true parameter is known. This was used in Section 4.3 to demonstrate that the proposed Laplacian chi-squared framework yields more accurate coverage intervals than a framework closer to more traditional approaches. The use of climate models in this validation study presents a valuable source of information for assessing the statistical properties of the methodology, however, the results from using a particular climate model do not necessarily guarantee the same statistical properties when applied to the observations due to potential climate model error. If the difference between

any particular climate model and the observed climate is assumed to be of similar magnitude as the difference between any two climate models from each other, then the scale of this error can be gauged by comparing the results obtained with different climate models. This can be seen in the validation study of Section 4.3, where distributions of results over pairs of historical and pre-industrial runs from different climate models are displayed.

The validation results show that the methods developed in this thesis advance the state of the art in their applied problems. Nevertheless, the results presented here reveal multiple promising avenues for future improvements and extensions. For the ocean heat content problem, the difference in the correlation structure between the January-only fit and the full spatio-temporal fit suggests that the correlation properties may themselves vary seasonally. Potential future work could focus on investigating the degree of statistical evidence for temporal non-stationarity in the parameter fields. Further insight could be gained by investigating potential physical reasons for changes in the correlation patterns either seasonally or over time.

On a broader level, the amount of heat being absorbed by the ocean is an important quantity for estimating equilibrium climate sensitivity, defined as the expected increase in global temperatures in response to a doubling of atmospheric carbon dioxide. The value of this parameter has immense practical importance for understanding the anticipated effects of anthropogenic climate change, however, the best estimate and the uncertainty in this estimate are notoriously difficult to quantify [Knutti et al., 2017]. Various work has been done in the climate literature to estimate and represent uncertainty in equilibrium climate sensitivity through Bayesian frameworks [Annan, 2015, Bodman and Jones, 2016, Goodwin and Cael, 2021]. Incorporating the Bayesian model for the trend in ocean heat content into these frameworks for climate sensitivity would be a promising direction for future research.

The Bayesian model for ocean heat content would also benefit from an extension of the data under consideration to additional portions of the ocean. In the applications of Chapter 2 and Chapter 3 only ocean heat content values integrated up to 2,000 meters of depth

within the mask displayed in Figure 2.1 were used. An even more accurate picture of ocean heat content could be obtained by extending the methodology to incorporate areas outside of the mask, heat content profiles with shallower maximum depths, and ocean heat content at deeper depths. In particular, the recent introduction of the deep Argo program [Jayne et al., 2017, Gasparin et al., 2020] allows for the analysis of ocean heat content up to 6,000 meters of depth in some areas of the ocean.

The Bayesian detection and attribution approach proposed in Chapter 4 prioritizes the representation of uncertainty in the covariance matrix, and as such does not incorporate several other important sources of uncertainty. In particular, observational uncertainty and uncertainty from estimating the forced signal have been major focuses of previous entries in the literature [Hammerling et al., 2019]. Such sources of uncertainty could be incorporated into the Bayesian framework presented here using established approaches. A more challenging source of uncertainty to incorporate is the uncertainty induced by climate model heterogeneity or error. One possibility for representing this uncertainty in the covariance matrix would be to add a level to the Bayesian hierarchy to "pool" the covariance parameters between different climate models. Such a direction would take advantage of fact that the Laplacian basis functions are the same between different models, a feature that is not present in principle component based approaches. Future research could also aim to incorporate heterogeneity in estimating the forced signal from historical runs with different climate models. As can be seen in the validation study of Section 4.3, these sources of uncertainty could be meaningful in magnitude and their representation would lead to more reliable detection and attribution results.

This proposed method is intended for the detection and attribution of climate changes at the global scale. While reliable detection and attribution methodology is important for emphasizing the benefit of reducing emissions, results from these methods have only a minor effect on the scientific consensus on the relationship between greenhouse gas concentrations and warming temperatures due to multiple supporting lines of evidence. However, establish-

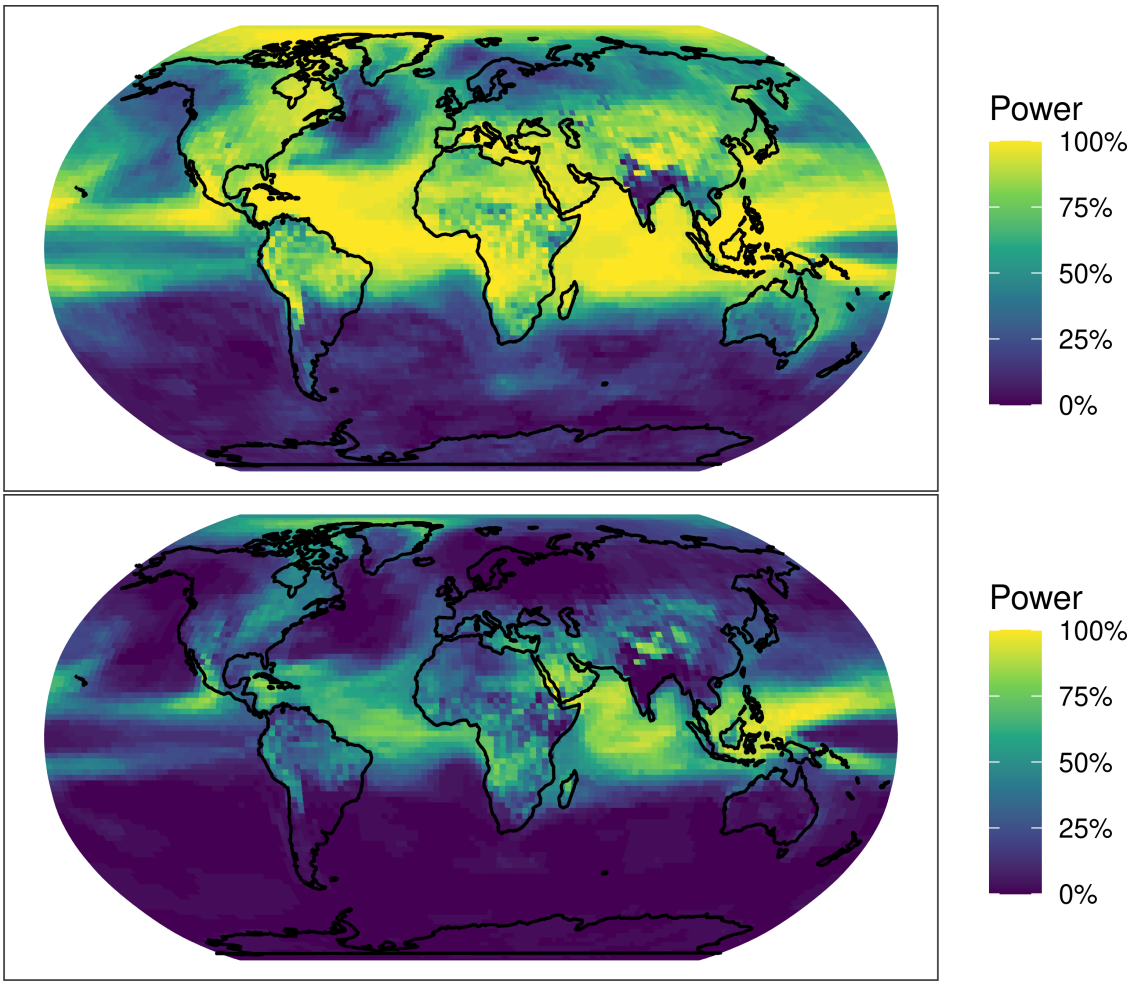


Figure 5.1: Power of the detection hypothesis test applied independently to each grid-cell without correcting for multiple testing (top panel) and with the application of the multiple testing correction of Benjamini and Yekutieli [2001] (bottom panel). Power is computed at the 5% level by averaging over the 47-member GISS ensemble [Miller et al., 2021].

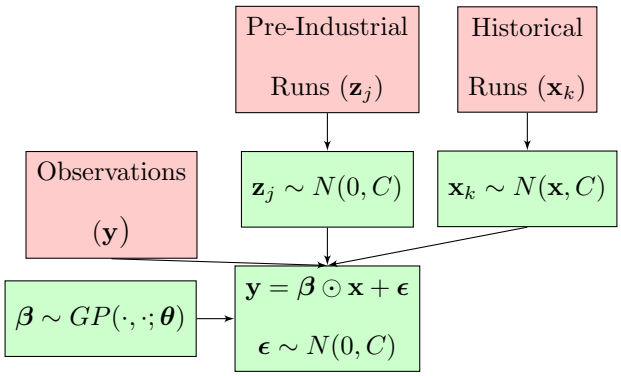


Figure 5.2: Diagram of a Bayesian hierarchical framework for representing detection and attribution parameters at the local level. Data is in red and distributional assumptions are in green. Indices j index over pre-industrial runs $\mathbf{z}_j$ and k index over historical runs $\mathbf{x}_k$. Here the symbol $\odot$ refers to pointwise vector multiplication. The vector $\boldsymbol{\theta}$ represents the parameters of the Gaussian process (GP) for $\boldsymbol{\beta}$.

ing the relationship between greenhouse gases and climate changes at smaller scales is more scientifically challenging. Furthermore, detection and attribution results for global trends are less relevant to policymakers and the public at large, who are generally more concerned with climate change impacts at their local level. Local detection and attribution approaches in the literature often employ statistical methods to each location independently [Knutson et al., 2013, Knutson and Zeng, 2018]. This runs into issues with the large number of hypotheses to be tested, and as demonstrated in Figure 5.1, correcting for multiple testing results in a substantial decrease in the power of successfully detecting the signal. This power could ostensibly be improved by taking into account the covariance structure describing the variability of the underlying climate process, as is done in the Bayesian detection and attribution framework for the global signal proposed in Chapter 4.

As such, extending the Bayesian detection and attribution framework to local scales is a highly motivated direction for future research. A potential approach would be to model the regression parameter as a spatially-varying field within the hierarchy. Then the value of this parameter field at any particular location would indicate the strength of the evidence

for detection and attribution while taking the global covariance structure into account. This idea, where the detection and attribution parameter is given a Gaussian process prior, is demonstrated in the diagram of Figure 5.2. The model in this diagram also incorporates observational uncertainty and uncertainty in the estimation of the forced signal as mentioned earlier in this section. The long-term goal of this line of research is to develop a Bayesian hierarchical model which takes into account all of the meaningful sources of uncertainty and can present the strength of evidence for a causal connection between climate forcings and effects at local scales.

The statistical methods developed in this thesis present quantitative improvements in their respective problems while illuminating promising directions for future research in the intersection of statistics and climate. Accurate quantification of uncertainty in climate problems is particularly important given the impacts of global warming on the environment, society, and daily life. As such, statistical methodology plays an important role in understanding the consequences of human activities on the climate. As greenhouse gas concentrations continue to increase, this role will become increasingly valuable in the years to come.

Bibliography

Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001.

Ron L Miller, Gavin A Schmidt, Larissa S Nazarenko, Susanne E Bauer, Maxwell Kelley, Reto Ruedy, Gary L Russell, Andrew S Ackerman, Igor Aleinov, Michael Bauer, et al. CMIP6 historical simulations (1850–2014) with GISS-E2. 1. *Journal of Advances in Modeling Earth Systems*, 13(1):e2019MS002034, 2021.

Sydney Levitus, John I. Antonov, Tim P. Boyer, Olga K. Baranova, Hernan Eduardo Garcia, Ricardo Alejandro Locarnini, Alexey V. Mishonov, James R. Reagan, Dan Seidov, Evgeney S. Yarosh, and Melissa M. Zweng. World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophysical Research Letters*, 39(10), 2012.

V Masson-Delmotte, P Zhai, A Priani, SL Connors, C Péan, S Berger, et al. IPCC, 2021: Climate Change 2021: The physical science basis. contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 2021.

Reto Knutti, Maria AA Rugenstein, and Gabriele C Hegerl. Beyond equilibrium climate sensitivity. *Nature Geoscience*, 10(10):727–736, 2017.

Svante Arrhenius. On the influence of carbonic acid in the air upon the temperature of the ground. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 41(251):237–276, 1896.

Jule G Charney, Akio Arakawa, D James Baker, Bert Bolin, Robert E Dickinson, Richard M Goody, Cecil E Leith, Henry M Stommel, and Carl I Wunsch. *Carbon dioxide and climate: a scientific assessment*. National Academy of Sciences, Washington, DC, 1979.

W Adger, S Agrawala, M Mirza, C Conde, K O'brien, J Pulhin, R Pulwarty, B Smit, and K Takahashi. Working group ii: impacts, adaptation and vulnerability. *Fourth assessment*

report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, UK, pages 717–743, 2007.

Peter M Caldwell, Mark D Zelinka, Karl E Taylor, and Kate Marvel. Quantifying the sources of intermodel spread in equilibrium climate sensitivity. *Journal of Climate*, 29(2):513–524, 2016.

Katarzyna B Tokarska, Gabriele C Hegerl, Andrew P Schurer, Piers M Forster, and Kate Marvel. Observational constraints on the effective climate sensitivity from the historical period. *Environmental Research Letters*, 15(3):034043, 2020.

Kevin E. Trenberth, John T. Fasullo, and Magdalena A. Balmaseda. Earth's energy imbalance. *Journal of Climate*, 27(9):3129–3144, 2014.

Reto Knutti and Lorenzo Tomassini. Constraints on the transient climate response from observed global temperature and ocean heat uptake. *Geophysical Research Letters*, 35(9), 2008.

Benoit Meyssignac, Tim Boyer, Zhongxiang Zhao, Maria Z. Hakuba, Felix W. Landerer, Detlef Stammer, Armin Köhl, Seiji Kato, Tristan L'Ecuyer, Michael Ablain, et al. Measuring global ocean heat content to estimate the earth energy imbalance. *Frontiers in Marine Science*, 6:432, 2019.

Lijing Cheng, Kevin E Trenberth, John Fasullo, Tim Boyer, John Abraham, and Jiang Zhu. Improved estimates of ocean heat content from 1960 to 2015. *Science Advances*, 3(3): e1601545, 2017a.

Laure Resplandy, Ralph F Keeling, Yassir Eddebbar, Mariela Brooks, Rao Wang, Laurent Bopp, Matthew C Long, John P Dunne, Wolfgang Koeve, and Andreas Oschlies. Quantification of ocean heat uptake from changes in atmospheric O2 and CO2 composition. *Scientific reports*, 9(1):1–10, 2019.

GC Hegerl, H Von Storch, K Hasselmann, BD Santer, U Cubasch, and PD Jones. Detecting anthropogenic climate change with an optimal fingerprint method. 1994.

Myles R Allen and Simon FB Tett. Checking for model consistency in optimal fingerprinting. *Climate Dynamics*, 15(6):419–434, 1999.

Dorit Hammerling, Matthias Katzfuss, and Richard Smith. Climate change detection and attribution. *Handbook of Environmental and Ecological Statistics. Chapman & Hall/CRC Handb. Mod. Stat. Methods*, pages 789–817, 2019.

Yan Li, Kun Chen, Jun Yan, and Xuebin Zhang. Uncertainty in optimal fingerprinting is underestimated. *Environmental Research Letters*, 16(8):084043, 2021.

Ross McKitrick. Checking for model consistency in optimal fingerprinting: a comment. *Climate Dynamics*, 58(1):405–411, 2022.

Nikolaos Christidis, Peter A Stott, Francis W Zwiers, Hideo Shiogama, and Toru Nozawa. The contribution of anthropogenic forcings to regional changes in temperature during the last decade. *Climate Dynamics*, 39(6):1259–1274, 2012.

Thomas R Knutson and Fanrong Zeng. Model assessment of observed precipitation trends over land regions: detectable human influences and possible low bias in model trends. *Journal of Climate*, 31(12):4617–4637, 2018.

Zhuo Wang, Yujing Jiang, Hui Wan, Jun Yan, and Xuebin Zhang. Toward optimal fingerprinting in detection and attribution of changes in climate extremes. *Journal of the American Statistical Association*, 116(533):1–13, 2021.

Julia Slingo and Tim Palmer. Uncertainty in weather and climate prediction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1956):4751–4767, 2011.

Stephen H Schneider and Robert E Dickinson. Climate modeling. *Reviews of Geophysics*, 12(3):447–493, 1974.

Warren M Washington and Claire Parkinson. *Introduction to three-dimensional climate modeling*. University science books, 2005.

Stephen Griffies. Fundamentals of ocean climate models. In *Fundamentals of Ocean Climate Models*. Princeton university press, 2018.

Daniel G Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6): 119–139, 1951.

Georges Matheron. Principles of geostatistics. *Economic geology*, 58(8):1246–1266, 1963.

Noel Cressie. The Origins of Kriging.-Mathematical Geology, vol. 22, pp. 239-252. 1990.

Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2003.

Michael L Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 1999a.

Noel Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.

David Higdon. A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, 5(2):173–190, 1998a.

Christopher J Paciorek and Mark J Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics: The official journal of the International Environmetrics Society*, 17(5):483–506, 2006a.

Mark D. Risser. Nonstationary spatial modeling, with emphasis on process convolution and covariate-driven approaches. *arXiv preprint arXiv:1610.02447*, 2016.

Jaehong Jeong and Mikyoung Jun. A class of matérn-like covariance functions for smooth processes on a sphere. *Spatial Statistics*, 11:1–18, 2015.

Yang Li and Zhengyuan Zhu. Modeling nonstationary covariance function with convolution on sphere. *Computational Statistics & Data Analysis*, 104:233–246, 2016a.

Joseph Guinness and Montserrat Fuentes. Isotropic covariance functions on spheres: Some properties and modeling considerations. *Journal of Multivariate Analysis*, 143:143–152, 2016.

Aldo V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):297–312, 1988.

Matthias Katzfuss, Joseph Guinness, Wenlong Gong, and Daniel Zilber. Vecchia approximations of Gaussian-process predictions. *Journal of Agricultural, Biological and Environmental Statistics*, 25(3):383–414, 2020a.

Joseph Guinness. Gaussian process learning via Fisher scoring of Vecchia's approximation. *Statistics and Computing*, 31(3):1–8, 2021.

Matthias Katzfuss and Joseph Guinness. A general framework for Vecchia approximations of Gaussian processes. *Statistical Science*, 36(1):124–141, 2021.

Daniel Zilber and Matthias Katzfuss. Vecchia–Laplace approximations of generalized Gaussian processes for big non-Gaussian spatial data. *Computational Statistics & Data Analysis*, 153:107081, 2021.

Felix Jimenez and Matthias Katzfuss. Scalable Bayesian optimization using Vecchia approximations of Gaussian processes. *arXiv preprint arXiv:2203.01459*, 2022.

Victor Minden, Anil Damle, Kenneth L Ho, and Lexing Ying. Fast spatial Gaussian process maximum likelihood estimation via skeletonization factorizations. *Multiscale Modeling & Simulation*, 15(4):1584–1611, 2017.

Samuel Baugh and Michael L Stein. Computationally efficient spatial modeling using recursive skeletonization factorizations. *Spatial Statistics*, 27:18–30, 2018.

Christopher J Paciorek. Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP package. *Journal of statistical software*, 19(2):nihpa22751, 2007.

Richard Courant and David Hilbert. *Methods of mathematical physics: partial differential equations*. John Wiley & Sons, 2008.

Timothy DelSole and Michael K Tippett. Laplacian eigenfunctions for climate analysis. *Journal of Climate*, 28(18):7420–7436, 2015.

L Mark Berliner, Richard A Levine, and Dennis J Shea. Bayesian climate change assessment. *Journal of Climate*, 13(21):3805–3820, 2000.

Richard L Smith, Claudia Tebaldi, Doug Nychka, and Linda O Mearns. Bayesian modeling of uncertainty in ensembles of climate models. *Journal of the American Statistical Association*, 104(485):97–116, 2009.

Michael E Mann, Elisabeth A Lloyd, and Naomi Oreskes. Assessing climate change impacts on extreme weather events: the case for an alternative (Bayesian) approach. *Climatic change*, 144(2):131–142, 2017.

Theodore G Shepherd. Bringing physical reasoning into statistical practice in climate-change science. *Climatic Change*, 169(1):1–19, 2021.

Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.

Stefan Van Dongen. Prior specification in Bayesian statistics: three cautionary tales. *Journal of theoretical biology*, 242(1):90–100, 2006.

James E Matheson and Robert L Winkler. Scoring rules for continuous probability distributions. *Management science*, 22(10):1087–1096, 1976.

Hans Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.

M Stone. Cross-validation: A review. *Statistics: A Journal of Theoretical and Applied Statistics*, 9(1):127–139, 1978.

Stephen Bates, Trevor Hastie, and Robert Tibshirani. Cross-validation: what does it estimate and how well does it do it? *arXiv preprint arXiv:2104.00673*, 2021.

Jennifer E Kay, Clara Deser, A Phillips, A Mai, Cecile Hannay, Gary Strand, Julie Michelle Arblaster, SC Bates, Gokhan Danabasoglu, James Edwards, et al. The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8):1333–1349, 2015.

Ryan L Sriver, Chris E Forest, and Klaus Keller. Effects of initial conditions uncertainty on regional climate variability: An analysis using a low-resolution CESM ensemble. *Geophysical Research Letters*, 42(13):5468–5476, 2015.

Brian C O'Neill, Claudia Tebaldi, Detlef P van Vuuren, Veronika Eyring, Pierre Friedlingstein, George Hurtt, Reto Knutti, Elmar Kriegler, Jean-Francois Lamarque, Jason Lowe, et al. The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9):3461–3482, 2016.

Michael L Stein. Some statistical issues in climate science. *Statistical Science*, 35(1):31–41, 2020.

Stephen C. Riser, Howard J. Freeland, Dean Roemmich, Susan Wijffels, Ariel Troisi, Mathieu Belbéoch, Denis Gilbert, Jianping Xu, Sylvie Pouliquen, Ann Thresher, Pierre-

Yves Le Traon, Guillaume Maze, Birgit Klein, M. Ravichandran, Fiona Grant, Pierre-Marie Poulain, Toshio Suga, Byunghwan Lim, Andreas Sterl, Philip Sutton, Kjell-Arne Mork, Pedro Joaquín Vélez-Belchí, Isabelle Ansorge, Brian King, Jon Turton, Molly Baringer, and Steven R. Jayne. Fifteen years of ocean observations with the global Argo array. *Nature Climate Change*, 6:145–153, January 2016. URL `https://doi.org/10.1038/nclimate2872`.

Samuel Baugh and Karen A McKinnon. Hierarchical Bayesian modeling of ocean heat content and its uncertainty. *Annals of Applied Statistics*, 2022.

Karina von Schuckmann and Pierre-Yves Le Traon. How well can we derive global ocean indicators from Argo data? *Ocean Science*, 7(6):783–791, 2011.

Viktor Gouretski. World ocean circulation experiment–Argo global hydrographic climatology. *Ocean Science*, 14(5):1127–1146, 2018.

Masayoshi Ishii, Yoshikazu Fukuda, Shoji Hirahara, Soichiro Yasui, Toru Suzuki, and Kanako Sato. Accuracy of global upper ocean heat content estimation expected from present observational data sets. *Sola*, 13:163–167, 2017.

Lijing Cheng and Jiang Zhu. Benefits of CMIP5 multimodel ensemble in reconstructing historical ocean subsurface temperature variations. *Journal of Climate*, 29(15):5393–5416, 2016.

Magdalena A. Balmaseda, Kevin E Trenberth, and Erland Källén. Distinctive climate signals in reanalysis of global ocean heat content. *Geophysical Research Letters*, 40(9):1754–1759, 2013.

Matthew D. Palmer, Chris D. Roberts, Magdalena A. Balmaseda, et al. Ocean heat content variability and change in an ensemble of ocean reanalyses. *Climate Dynamics*, 49(3): 909–930, 2017.

Michael L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999b.

Mikael Kuusela and Michael L. Stein. Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2220):20180400, 2018. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2018.0400.

Lijing Cheng, Kevin E. Trenberth, John Fasullo, Tim Boyer, John Abraham, and Jiang Zhu. Improved estimates of ocean heat content from 1960 to 2015. *Science Advances*, 3(3), 2017b. URL https://advances.sciencemag.org/content/3/3/e1601545.

Henk A. Dijkstra. *Dynamical Oceanography*. Springer, November 2008. ISBN 3540763759. URL https://www.xarg.org/ref/a/3540763759/.

Jaehong Jeong, Mikyoung Jun, and Marc G. Genton. Spherical process models for global spatial statistics. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 32(4):501, 2017.

David Higdon. A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*, 5(2):173–190, 1998b.

Christopher J. Paciorek and Mark J. Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5):483–506, 2006b. ISSN 1180-4009. URL https://www.ncbi.nlm.nih.gov/pubmed/18163157.

Yang Li and Zhengyuan Zhu. Modeling nonstationary covariance function with convolution on sphere. *Computational Statistics and Data Analysis*, 104:233–246, 2016b. ISSN 0167-9473. URL http://www.sciencedirect.com/science/article/pii/S0167947316301566.

Paul D. Sampson and Peter Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119, 1992.

Marcin Hitczenko and Michael L. Stein. Some theory for anisotropic processes on the sphere. *Statistical Methodology*, 9:211–227, 03 2012.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

Daniel Turek and Mark D. Risser. BayesNSGP: Bayesian Analysis of Non-Stationary Gaussian Process Models, October 2019. URL `https://CRAN.R-project.org/package=BayesNSGP`.

Mark D. Risser and Daniel Turek. Bayesian inference for high-dimensional nonstationary Gaussian processes. *Journal of Statistical Computation and Simulation*, 90(16):2902–2928, November 2020. ISSN 0094-9655.

Mark Risser. Nonstationary Bayesian modeling for a large data set of derived surface temperature return values. *arXiv:2005.03658 [stat]*, April 2020. URL `http://arxiv.org/abs/2005.03658`.

Dean Roemmich and John Gilson. The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo program. *Progress in Oceanography*, 82(2):81 – 100, 2009. ISSN 0079-6611. URL `http://www.sciencedirect.com/science/article/pii/S0079661109000160`.

Angela Colling. *Ocean circulation*, volume 3. Butterworth-Heinemann, 2001.

Matthias Katzfuss. Bayesian nonstationary spatial modeling for very large datasets. *Environmetrics*, 24(3):189–200, 2013.

Gareth O. Roberts and Jeffrey S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

Joseph Guinness. Permutation and grouping methods for sharpening gaussian process approximations. *Technometrics*, 60(4):415–429, 2018.

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

Matthias Katzfuss, Joseph Guinness, Wenlong Gong, and Daniel Zilber. Vecchia approximations of Gaussian-process predictions. *Journal of Agricultural, Biological and Environmental Statistics*, 25(3):383–414, 2020b.

Philip Heidelberger and Peter D. Welch. A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM*, 24(4):233–245, 1981.

Tilmann Gneiting, Adrian E. Raftery, Anton H. Westveld III, and Tom Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133(5):1098–1118, 2005.

Jopseh Guinness and Matthias Katzfuss. GpGp: Fast Gaussian process computation using Vecchia's approximation. *R package version 0.1. 0*, 2018.

Matthias Katzfuss, Marcin Jurek, Daniel Zilber, Wenlong Gong, Joseph Guinness, Jingjie Zhang, and Florian Schäfer. GPvecchia: Fast Gaussian-process inference using Vecchia approximations. *R package version 0.1*, 3, 2020c.

Michael H Glantz, Michael H Glantz, et al. *Currents of change: impacts of El Nino and La Nina on climate and society*. Cambridge University Press, 2001.

Lijing Cheng, John Abraham, Zeke Hausfather, and Kevin E Trenberth. How fast are the oceans warming? *Science*, 363(6423):128–129, 2019.

Gabriele C Hegerl, Hans von Storch, Klaus Hasselmann, Benjamin D Santer, Ulrich Cubasch, and Philip D Jones. Detecting greenhouse-gas-induced climate change with an optimal fingerprint method. *Journal of Climate*, 9(10):2281–2306, 1996.

Myles R Allen and Peter A Stott. Estimating signal amplitudes in optimal fingerprinting, Part I: Theory. *Climate Dynamics*, 21(5-6):477–491, 2003.

Chris Huntingford, Peter A Stott, Myles R Allen, and F Hugo Lambert. Incorporating model uncertainty into attribution of observed temperature change. *Geophysical Research Letters*, 33(5), 2006.

Alexis Hannart, Aurélien Ribes, and Philippe Naveau. Optimal fingerprinting under multiple sources of uncertainty. *Geophysical Research Letters*, 41(4):1261–1268, 2014.

Aurélien Ribes, Jean-Marc Azaïs, and Serge Planton. Adaptation of the optimal fingerprint method for climate change detection using a well-conditioned covariance matrix estimate. *Climate Dynamics*, 33(5):707–722, 2009.

Aurélien Ribes, Serge Planton, and Laurent Terray. Application of regularised optimal fingerprinting to attribution. Part I: method, properties and idealised analysis. *Climate dynamics*, 41(11):2817–2836, 2012.

Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

Alexis Hannart. Integrated optimal fingerprinting: Method description and illustration. *Journal of Climate*, 29(6):1977–1998, 2016.

Aurélien Ribes, Francis W Zwiers, Jean-Marc Azais, and Philippe Naveau. A new statistical approach to climate change detection and attribution. *Climate Dynamics*, 48(1):367–386, 2017.

Matthias Katzfuss, Dorit Hammerling, and Richard L Smith. A Bayesian hierarchical model for climate change detection and attribution. *Geophysical Research Letters*, 44(11):5720–5728, 2017.

Jennifer A Hoeting, David Madigan, Adrian E Raftery, and Chris T Volinsky. Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and EI George, and a rejoinder by the authors. *Statistical science*, 14(4):382–417, 1999.

Timothy DelSole, Laurie Trenary, Xiaoqin Yan, and Michael K Tippett. Confidence intervals in optimal fingerprinting. *Climate Dynamics*, 52(7):4111–4126, 2019.

F Lambert, P Stott, and M Allen. Detection and attribution of changes in global terrestrial precipitation. In *EGS-AGU-EUG Joint Assembly*, page 6140, 2003.

Xuebin Zhang, Francis W Zwiers, Gabriele C Hegerl, F Hugo Lambert, Nathan P Gillett, Susan Solomon, Peter A Stott, and Toru Nozawa. Detection of human influence on twentieth-century precipitation trends. *Nature*, 448(7152):461–465, 2007.

Seung-Ki Min, Xuebin Zhang, Francis W Zwiers, and Gabriele C Hegerl. Human contribution to more-intense precipitation extremes. *Nature*, 470(7334):378, 2011. doi: 10.1038/nature09763.

Hui Wan, Xuebin Zhang, Francis Zwiers, and Seung-Ki Min. Attributing northern high-latitude precipitation change over the period 1966–2005 to human influence. *Climate Dynamics*, 45(7):1713–1726, 2015.

Ben D Santer, Michael F Wehner, TML Wigley, R Sausen, GA Meehl, KE Taylor, C Am-

mann, J Arblaster, WM Washington, JS Boyle, et al. Contributions of anthropogenic and natural forcing to recent tropopause height changes. *science*, 301(5632):479–483, 2003.

Benjamin D Santer, Jeffrey F Painter, Carl A Mears, Charles Doutriaux, Peter Caldwell, Julie M Arblaster, Philip J Cameron-Smith, Nathan P Gillett, Peter J Gleckler, John Lanzante, et al. Identifying human influences on atmospheric temperature. *Proceedings of the National Academy of Sciences*, 110(1):26–33, 2013.

Benjamin D Santer, Stephen Po-Chedley, Mark D Zelinka, Ivana Cvijanovic, Céline Bonfils, Paul J Durack, Qiang Fu, Jeffrey Kiehl, Carl Mears, Jeffrey Painter, et al. Human influence on the seasonal cycle of tropospheric temperature. *Science*, 361(6399):eaas8806, 2018.

Shuangmei Ma, Tianjun Zhou, Dáithí A Stone, Debbie Polson, Aiguo Dai, Peter A Stott, Hans von Storch, Yun Qian, Claire Burke, Peili Wu, et al. Detectable anthropogenic shift toward heavy precipitation over eastern China. *Journal of Climate*, 30(4):1381–1396, 2017.

Jingpeng Zhang, Tianbao Zhao, Aiguo Dai, and Wenyu Zhang. Detection and attribution of atmospheric precipitable water changes since the 1970s over China. *Scientific Reports*, 9(1):1–10, 2019.

OD Andrews, NL Bindoff, PR Halloran, Tatiana Ilyina, and Corinne Le Quéré. Detecting an external influence on recent changes in oceanic oxygen using an optimal fingerprinting method. *Biogeosciences*, 10(3):1799–1813, 2013.

Hugo G Hidalgo, Tapash Das, Michael D Dettinger, Daniel R Cayan, David W Pierce, Tim P Barnett, Govindasamy Bala, Aq Mirin, Andrew W Wood, Celine Bonfils, et al. Detection and attribution of streamflow timing changes to climate change in the western United States. *Journal of Climate*, 22(13):3838–3855, 2009.

Jiafu Mao, Aurélien Ribes, Binyan Yan, Xiaoying Shi, Peter E Thornton, Roland Séférian, Philippe Ciais, Ranga B Myneni, Hervé Douville, Shilong Piao, et al. Human-induced

greening of the northern extratropical land surface. *Nature Climate Change*, 6(10):959–963, 2016.

Nikolaos Christidis, Peter A Stott, Gabriele C Hegerl, and Richard A Betts. The role of land use change in the recent warming of daily extreme temperatures. *Geophysical Research Letters*, 40(3):589–594, 2013.

Yeon-Hee Kim, Seung-Ki Min, Xuebin Zhang, Francis Zwiers, Lisa V Alexander, Markus G Donat, and Yu-Shiang Tung. Attribution of extreme temperature changes during 1951–2010. *Climate dynamics*, 46(5):1769–1782, 2016.

Chunhui Lu, Ying Sun, Hui Wan, Xuebin Zhang, and Hong Yin. Anthropogenic influence on the frequency of extreme temperatures in China. *Geophysical Research Letters*, 43(12): 6511–6518, 2016.

Min-Gyu Seong, Seung-Ki Min, Yeon-Hee Kim, Xuebin Zhang, and Ying Sun. Anthropogenic greenhouse gas and aerosol contributions to extreme temperature changes during 1951–2015. *Journal of Climate*, 34(3):857–870, 2021.

Axel Timmermann. Detecting the nonstationary response of ENSO to greenhouse warming. *Journal of the Atmospheric Sciences*, 56(14):2313–2325, 1999.

Christopher D Roberts and Matthew D Palmer. Detectability of changes to the Atlantic meridional overturning circulation in the Hadley Centre Climate Models. *Climate dynamics*, 39(9):2533–2546, 2012.

William Richard Hobbs, Nathaniel L Bindoff, and Marilyn N Raphael. New perspectives on observed and simulated antarctic sea ice extent trends using optimal fingerprinting techniques. *Journal of Climate*, 28(4):1543–1560, 2015.

Evan Weller, Seung-Ki Min, Wenju Cai, Francis W Zwiers, Yeon-Hee Kim, and Donghyun

Lee. Human-caused Indo-Pacific warm pool expansion. *Science advances*, 2(7):e1501719, 2016.

Colin P Morice, John J Kennedy, Nick A Rayner, JP Winn, Emma Hogan, RE Killick, RJH Dunn, TJ Osborn, PD Jones, and IR Simpson. An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres*, 126(3):e2019JD032361, 2021.

Alexander C Aitken. Iv.—on least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1936.

Calyampudi Radhakrishna Rao. *Linear statistical inference and its applications*, volume 2. Wiley New York, 1973.

Veronika Eyring, Sandrine Bony, Gerald A Meehl, Catherine A Senior, Bjorn Stevens, Ronald J Stouffer, and Karl E Taylor. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016.

Timothy J Osborn, Philip D Jones, David H Lister, Colin P Morice, Ian R Simpson, JP Winn, Emma Hogan, and Ian C Harris. Land surface air temperature variations across the globe updated to 2019: The CRUTEM5 data set. *Journal of Geophysical Research: Atmospheres*, 126(2):e2019JD032352, 2021.

Brian Kidd and Matthias Katzfuss. Bayesian nonstationary and nonparametric covariance estimation for large spatial data. *Bayesian Analysis*, 1(1), 2021.

James Annan. Recent developments in Bayesian estimation of climate sensitivity. *Current Climate Change Reports*, 1(4):263–267, 2015.

Roger W Bodman and Roger N Jones. Bayesian estimation of climate sensitivity using ob-

servationally constrained simple climate models. *Wiley Interdisciplinary Reviews: Climate Change*, 7(3):461–473, 2016.

Philip Goodwin and BB Cael. Bayesian estimation of earth's climate sensitivity and transient climate response from observational warming and heat content datasets. *Earth System Dynamics*, 12(2):709–723, 2021.

Steven R Jayne, Dean Roemmich, Nathalie Zilberman, Stephen C Riser, Kenneth S Johnson, Gregory C Johnson, and Stephen R Piotrowicz. The Argo program: present and future. *Oceanography*, 30(2):18–28, 2017.

Florent Gasparin, Mathieu Hamon, Elisabeth Rémy, and Pierre-Yves Le Traon. How deep Argo will improve the deep ocean in an ocean reanalysis. *Journal of Climate*, 33(1):77–94, 2020.

Thomas R Knutson, Fanrong Zeng, and Andrew T Wittenberg. Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 twentieth-century simulations. *Journal of Climate*, 26(22):8709–8743, 2013.