HIGH-DIMENSIONAL MANIFOLD GEOSTATISTICS

By

CHINTAN ALAKH DALAL

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

Dimitris N. Metaxas

And approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2019

ABSTRACT OF THE DISSERTATION

High-Dimensional Manifold Geostatistics

By CHINTAN ALAKH DALAL

Dissertation Director:

Dimitris N. Metaxas

An understanding of the Earth's climate system benefits all sectors of the economy and environment. Several challenges faced when modeling the Earth's climate system include: estimating geographical features of global datasets, making inferences from multiple data-products, and providing diagnostic tools for complex Earth models. Existing geostatistical approaches address these challenges by modeling points on a high-dimensional space. However, we know that many of the climate datasets additionally have inherent high-dimensional geometric structures.

In this dissertation, I provide new insights into problems in climate data science by exploring high-dimensional geometric structures on a manifold. First, I will discuss an approach to improve future projections of a climate variable (e.g., sea-level changes) by learning the scale of correlation, an essential regional feature of climate datasets. Second, I will provide a new framework for data-fusion from multiple sources of information for a given climate variable. Third, I will describe diagnostic tools we created to compare and emulate various Earth system models from numerous international teams and for differing future climate scenarios (e.g. precipitation changes in 2090). With these contributions, I demonstrate that we can improve the inferences made from geostatistical models by including information about the high-dimensional structures of climate datasets. The proposed novel framework will benefit not only the climate community but also decision makers when identifying plans to mitigate the impact of climate change.

Acknowledgements

This work represents the contributions of my family, friends, colleagues, mentors, and institutions for which I am grateful.

First and foremost, I am grateful to my advisor, Dimitris Metaxas, for his continuous support and encouragement throughout the various phases of my academic journey, from robotics to computer vision, to biomedical imaging, to differential geometry, and, finally, to climate science. Additionally, I want to extend my gratitude to Douglas Nychka for enlightening me in the field of statistics and data science. This work wouldn't have come to a culmination without his guidance and stimulating discussions.

A very special thanks goes to Vladimir Pavlovic who introduced me to the two fields that this dissertation is primarily concerned with, that is information geometry and climate science. His insightful comments and hard questions have helped in pushing my intellect towards concretizing this work. Also, many thanks to my dissertation committee members Jonathan Stroud and Kostas Bekris for going out of their way to support this work in its critical moments.

I am also indebted to Robert Kopp, for his infectious passion towards the problems in climate science. My inspiration to continue applying unorthodox approaches to climate science delves from the many interesting conversations we have had at various coffee shops around New Jersey over the last four years. Thanks also for introducing me to Colorado and the National Center for the Atmospheric Research (NCAR). During my unending internship at NCAR, I had a great pleasure working with Claudia Tebaldi, who I would like to thank for introducing me to the many exciting challenges in Bayesian statistics and its application to climate science.

Over the many years of my graduate studies, I feel humbled to have met pioneering individuals who have helped me to move past challenges when most required. In

particular, I am grateful to Arindam Banerjee for believing in my work and reassuring its importance. Thanks also to Michael Stein and Christopher Paciorek for providing constructive remarks and alternative viewpoints on my approach to geostatistical models. I further extend my thanks to George Moustakides and Susovan Pal for patiently reviewing and probing my work related to information geometry.

My view of the world as yet another concept in geometry stems from my formative years in robotics working with Daniel Lee and Vijay Kumar at the University of Pennsylvania, and, later, with Barry Bodt and Minbo Shim at General Dynamics. I thank each one of them for providing me with the essential tools of how to think and conduct academic research. More so, I am thankful to have worked alongside Daniel Lee, as that experience serves me to this day with the ideals of original thinking.

The freedom to think in my graduate studies would not have been possible without the support of (CCICADA) Command, Control, and Interoperability Center for Advanced Data Analysis (under Grant 2012-ST-104-000044). At CCICADA, I would especially like to thank Tamra Carpenter who nurtured my professional growth and provided valuable advice for improving my graduate experience. Thanks also to the silent contributors, Carol DiFrancesco at Rutgers and Silvia Gentile at NCAR, for smoothing out the many operational obstacles and aiding in my wonderful graduate experience.

I want to acknowledge all of my colleagues with whom I have cherished the many stimulating conversations and adventures. Special thanks to Matthew Edwards for painstakingly reviewing this dissertation and for always being enthusiastic about exploring wildly unrealistic ideas in statistics. I also extend my thanks to Chetan Tonde, Danna Gurari, Chao Chen, Jingjing Liu, and Felipe Tagle for providing many rounds of useful feedback throughout my doctoral research work.

I feel fortunate to have been surrounded by wonderful friends: Anatoliy Kats, Jonathan Williford, Abhishek Surana, Mubeen Bhatti, Ying Zheng, and Carol Raehn. Thank you all for your warmth during the different stages of my life. Especially, thanks to Shobhit Verma for his spiritual guidance in the practice of attaining sudden unenlightenment. Moving away from unenlightenment, I thank Shaula Gurari for her nudge into an epicurious existence filled with gourmet delights.

Most importantly, my heartfelt regards goes to all of my family both far and near. Thanks to my aunts Varsha, Vandana, and Raksha for the constant source of sweet words and foods. I also wish to thank my grandparents Dhanksukhlal and Bhadraben for raising me with the many Indian superhero fables, as well as for the bandages after my soccer practice. Thank you for rightly shaping my being. I owe a special thanks to my friend, world travel companion, long-distance lover, dissertation reviewer, and last but not the least my wife – Netta! Thank you for your unwavering patience and love. I salute you.

My deepest gratitude goes to my brother, Bhavin, and my parents, Asha and Alakh. Thanks to Bhavin for always having my back, for the many math and science tutorials, and for leading through your exemplary life. Finally, I believe that no words could express my feelings in gratitude towards my parents Asha and Alakh Dalal, and that only silence could...

Dedication

I dedicate this work to my family, Alakh, Asha, Bhavin, Bhadraben, Netta, and Ruan, for their unconditional love and smiles.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

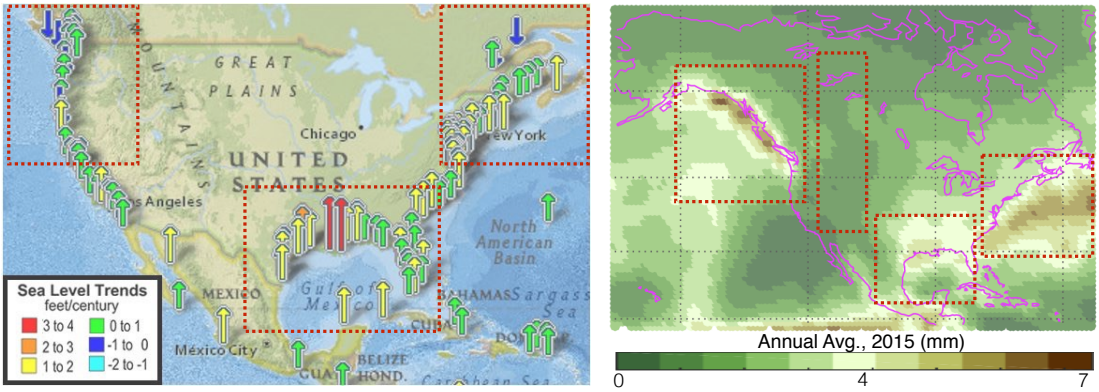
1.1 Overview

Geostatistics is a branch of statistics that deals with estimation and simulation of continuous spatial phenomena from a sample of measurements [1, 2, 3]. Geostatistics finds its use in various disciplines such as climate, epidemiology, environment, ecology, and astronomy [4, 5]. In this thesis, our motivation for the geostatistical model development is driven by its application to the climate data science problems of estimating sea-level changes and simulating future scenarios of precipitation changes.

To study problems in climate data science, in this era of large datasets, one has the possibility of exploiting high-dimensional geometric structures from the various sources of datasets [5, 6]. Figure 1.1 gives an example of one such geometric structure that arises from the regional properties of the natural phenomena. In the geostatistical literature, modeling these structures using anisotropy and a spatially-varying non-stationary Gaussian process has shown promise [1, 7, 8].

Albeit, until now, the climate data structures mentioned above have not been studied in its intrinsic metric space, specifically, the tangent spaces. Analyzing geometric structures on tangent spaces have recently shown success in applications where the data is represented as structure tensors, such as in computer vision and bio-medical imaging [9, 10, 11].

Our essential insight for developing geostatistical models is the property of the model's core elements, such as covariance matrix and anisotropy, which is symmetric positive definite (SPD). SPD space induces an intrinsic metric, i.e., a Riemannian metric, on a manifold that provides us with the theoretical foundation for developing the



(a) (b)

Figure 1.1: Examples of regional geophysical phenomena are depicted within the red boxes. We propose to represent regional climate data structures as a high-dimensional geometric object on a manifold. The datasets shown here are from (a) tide-gauge sites and (b) a precipitation data-product around North America.

high-dimensional manifold geostatistical models.

The two independently developed fields in statistics that this work is primarily concerned with are: intrinsic statistics for sample points that lie on a manifold and geostatistical models for sample points that lie on a globe.

Rao [12] in 1949 initially introduced the concept of intrinsic statistics for geometric objects, such as probability distributions of exponential families. Around the same time, Krige [13] in 1951 and Matheron [14] in 1963 developed the principles of geostatistics for spatial predictions in mining operations.

For our methodological development, we propose to synthesize the two fields in statistics mentioned above to study scientific problems in climate. Specifically, by providing an extra layer of inference on the high-dimensional manifold, we aid the classical geostatistical model development. The geostatistical model we focus on is that of the non-stationary Gaussian process and multi-variate normal sampling scheme, for climate datasets.

To begin, we explore and extract the high-dimensional information from the sea-level change datasets in spatial $m = 2$ and spatiotemporal $m = 3$ dimensions (e.g., see Figure 1.1a). By exploring the high-dimensional information, we show improvements in estimating the parameters of a geostatistical model for the non-stationary process.

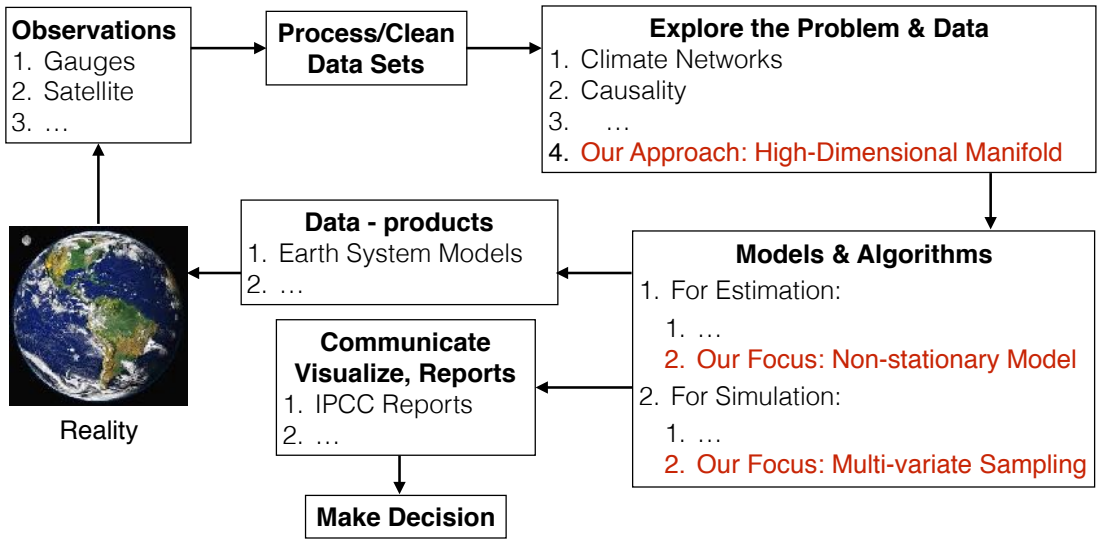


Figure 1.2: Overview of climate data science cycle. The red text within the boxes emphasize the focus of this thesis.

For higher dimensions, i.e., $m \geq 50$, we explore the Earth system model's outputs of precipitation changes (e.g., see Figure 1.3, bottom row). Subsequently, by using high-dimensional information, we show improvements in simulating Earth's future climate scenarios from a multi-variate normal sampling scheme.

Figure 1.2 places the focus of this thesis in an emerging field of climate data science [15, 16, 17].

1.2 Geostatistics: Problem Statement

The most commonly used geostatistical model for understanding Earth's underlying physical process for climate and weather prediction assumes a real-valued stochastic process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{G} \subset \mathbb{R}^{\text{dim}}\}$ that is spatially continuous, and many-times differentiable, over the globe $\mathcal{G}$. In practice, we only have finite collection of observations, $Y(\boldsymbol{s}) = \{Y(\boldsymbol{s_1}), \ldots, Y(\boldsymbol{s_n})\}$, at geolocations, $\boldsymbol{s} = \{\boldsymbol{s_1}, \ldots, \boldsymbol{s_n}\}$, where $n = (1, 2, \ldots)$ and $\boldsymbol{s}_i \in \mathbb{R}^{\text{dim}}$. Then the canonical objective is to make inference about the underlying process, $Z(\cdot)$, from the observations, $Y(\boldsymbol{s})$.

For notations let the covariate (or feature) space be denoted as $\boldsymbol{x}(\boldsymbol{s_i}) \in \mathbb{R}^m$, where $\boldsymbol{x}(\boldsymbol{s_i}) = \{x^1(\boldsymbol{s_i}), \ldots, x^m(\boldsymbol{s_i})\}$. For example, in the climate-context the covariate space

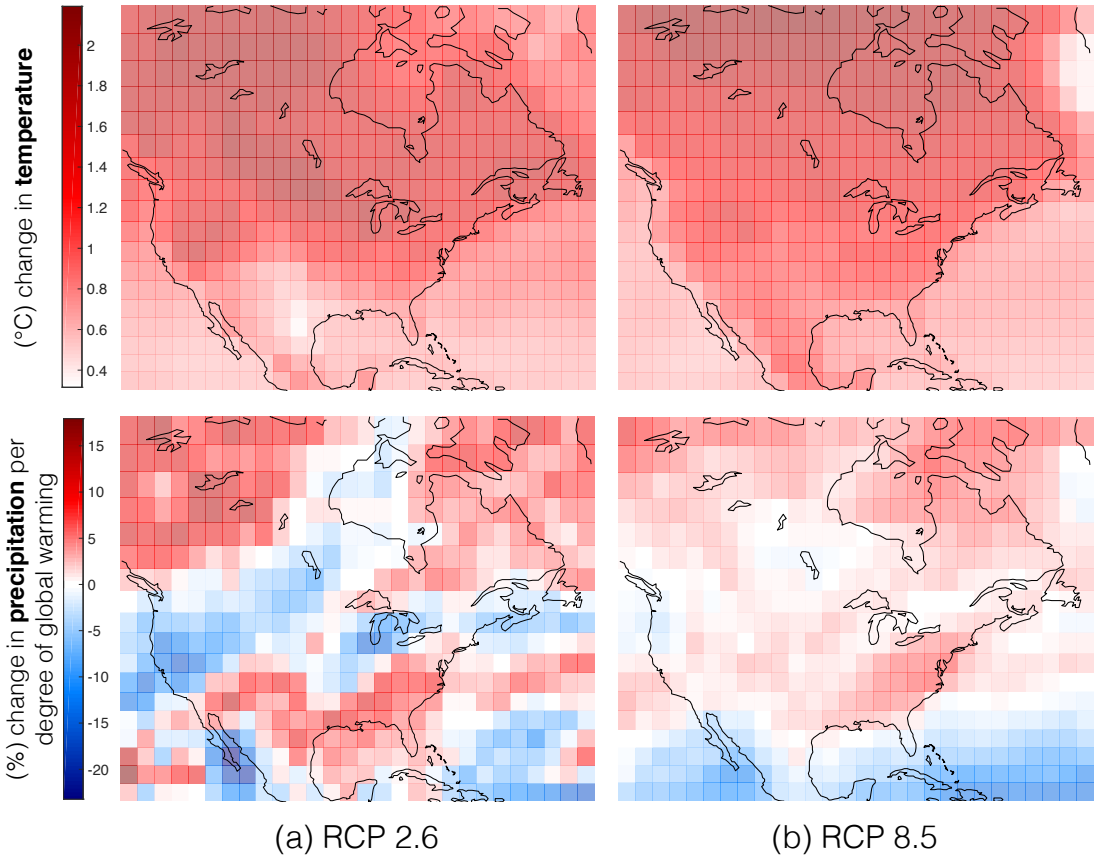


Figure 1.3: End of the 21st-century projections in temperature (top row) and precipitation (bottom row) around North America from the Community Climate System Model (CCSM version 4.0). The CCSM4 model is developed at the National Center for Atmospheric Research, USA. The units for temperature are °C and for precipitation are % change per °C change in global mean temperature. The columns show two scenarios: (a) representative concentration pathways (RCP) 2.6, i.e., when the greenhouse gas (GHG) emission is sustainably controlled, and (b) RCP 8.5, i.e., when the GHG emission is not controlled.

can be viewed as $\{\text{latitude, longitude, elevation, time}\}_i$. We also refer to $\mathbb{R}^m$ as the high-dimensional covariate space, and, sometimes interchangeable use it to denote $\mathbb{R}^{\text{dim}}$ when the context is clear.

Next, we briefly describe the two geostatistical inference problems of estimation and simulation where we explore our high-dimensional manifold approach.

Geostatiscal model for estimation

To learn the underlying stochastic model from the data, an important ingredient is the covariance function, $k(\boldsymbol{s}, \boldsymbol{s}') = Cov\{Z(\boldsymbol{s}), Z(\boldsymbol{s}')\}$. The covariance measures the

relation between two stochastically independent geolocations $\boldsymbol{s}$ and $\boldsymbol{s}'$. For example, the covariance k(New York, Florida) tells us how the sea-level changes around New York would affect the changes in Florida, and vice-versa. Our goal is to design the covariance structure that captures this complex relationship of physical processes from the available datasets.

Most physical and dynamical systems have a covariance structure that is anisotropic and non-stationary [5]. Intuitively, a non-stationary process means that the parameters of the underlying distribution, such as the mean and variance, varies spatiotemporally. Similarly, the anisotropic process, tells us that the properties of the process are directionally dependent. The most straightforward case of anisotropy, and the one we choose to model, i.e., geometric anisotropy, relates to the differing range of correlation between geolocations for each input space axis.

Spatial non-stationary covariance with geometric anisotropy has been widely studied in the field of geostatistics [5]. The two primary schools of thought are that of deformation models and spatially varying models [18]. Deformation model transforms the input space to a space that is stationary and isotropic [19], while spatially varying model convolves the kernel at each location with Brownian noise [7, 8].

Both of the above mentioned spatial non-stationary models have computational limitations for the parameter estimation in high-dimensional covariate space. Additionally, these methods, have only considered the correlation between geolocations and not between the regional structure of the underlying process from single and multiple sources of datasets. We show how the Riemannian manifold representation improves the aforementioned design limitations of the non-stationary covariance structure, and thereby improves the estimate of the sea-level changes in spatial $m = 2$ and spatiotemporal $m = 3$ dimensions.

For estimation and prediction, the standard non-parametric regression model that is extensively applicable to physical processes is the Gaussian process model (GP), also known as Kriging in geostatistics. One of the reasons GP is popular is because it can be completely specified by its first and second-order moments, i.e., mean and covariance functions. We refer to [3] for GP's in-depth theoretical exposition, and [1]

for its methodological applicability to physical processes.

Geostatiscal model for simulation

We now turn our attention to the geostatistical sampling problem to show the applicability of the manifold representation of high-dimensional ($m \geq 50$) covariance structures for physical processes. Specifically, we study a significant problem in climate science, which is that of simulating various plausible future climate scenarios from the existing ensemble of climate model outputs [20, 21]. For example, Figure 1.3 shows the future climate scenarios of temperature and precipitation around North America from an Earth system model (ESM).

For simulating the spatial field, $\{\boldsymbol{Z}(\boldsymbol{s})\}$, we focus on improving the vanilla multivariate normal sampling scheme [4, 22, 23]. This straightforward sampling scheme uses an affine transformation $\boldsymbol{Z} = \boldsymbol{\mu} + A\boldsymbol{\epsilon}$, where $\boldsymbol{Z} = (\boldsymbol{Z}(\boldsymbol{s_1}), \ldots, \boldsymbol{Z}(\boldsymbol{s_m}))^T$ is a m-dimensional (multi-variate) vector, A is the affine matrix such that $AA^T = \Sigma$, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a random normal vector. Furthermore, this sampling scheme is a basic component for a more widely used framework of hierarchical Bayesian models [6].

In the geostatistics literature, parameters of a multi-variate normal sampling scheme, such as mean and covariance, are estimated by assuming independence between the replicates of the observations. In our application, the replicates of climate model outputs are known to have dependencies. For example, the multiple images in Figure 1.3 are known to have various inter-dependencies. Additionally, the vanilla sampling scheme and its extensions to the hierarchical Bayesian model fails to embed spatial patterns between the different climate model outputs [24, 25].

To address the above sampling issues, we represent the initial estimates of the covariance structure on a Riemannian manifold of symmetric positive definite matrices, $\Sigma \in \mathrm{SPD}(m)$, of dimension m. This reformulation provides not just an unbiased estimator, but also a distance function to compare the dependencies between various sources of the datasets, which in our case are the various climate model outputs.

Our primary motivation for the geostatistical model development is that of improving the understanding of climate-related problems. In the next section, we briefly

describe the four scientific problems where we apply our proposed high-dimensional manifold geostatistical models.

1.3 Application of Geostatistics: Problems in Climate Science

The 20th-century global mean sea level (GMSL) is estimated to have increased at a mean rate of 1.6 to 1.9mm/year [26, 20]. Additionally, from 1990 to 2010, the GMSL rate of change has risen to 3.0 $\pm$ 1.7mm/year [27, 28]. Due to the implications of sea-level changes, it is projected that 153 million lives are going to be affected by the end of this century [29].

Estimating regional sea-level changes

Estimating the impact and extreme events due to regional sea-level changes (ReSL) can be far more significant than the global mean, i.e., GMSL, [30, 31]. Additionally, ReSL can have a large variability compared to GMSL due to its complex spatial patterns resulting from ocean dynamics, sea-floor movements, and water mass redistribution [20, 32, 33].

Most climate science approaches rely on the known geophysical and geographical literature to provide ReSL estimates [34, 31, 35]. In contrast, we provide a systematic data-driven approach to incorporate spatial structures for estimating ReSL.

To show the applicability of the proposed framework for estimating ReSL we use the data-products of tide-gauge records and satellite altimeter, around North America. The tide-gauge records measure the sea-level changes at coastal locations around the global network, while the satellite altimeter measures the spatially dense sea surface height over the oceans.

Moreover, the data-product of tide-gauge records is provided by the permanent service for mean sea level (PSMSL) [36], and the satellite altimeter datasets are from the TOPEX and Jason satellites [37, 38].

Inference from multiple sources of datasets

With the advent of the data era and remote sensing technology, climate community has become increasingly interested in utilizing measurements from various sources to explain the climate-related processes. For example, one question addressed in our work is how to use both the tide gauges and satellite altimeter to improve our understanding of sea-level changes.

Both tide gauges and satellite altimeter provides measurements at different spatiotemporal scales. Compared to satellite altimeter measurements the tide-gauge measurements provide more information about the historical sea levels. Even so, tide-gauge measurements are spatial very sparse when compared to satellite altimeter measurements.

Chambers utilized measurements from multiple data sources to estimate GMSL using a standard principal component analysis (PCA) [39]. Note, in the climate community, PCA is commonly known as empirical orthogonal functions (EOF). Recently, [34, 26, 28] have improved GMSL estimates by utilizing both tide-gauges and satellite altimeter. For ReSL, [30] provides an EOF-based methodology to improve regional estimates.

Limitations of standard PCA (i.e., EOF) are studied widely in the machine learning community, and many variants of the standard PCA techniques are continually advancing. Many of the advanced PCA techniques also incorporate the non-linearity and intrinsic metric of the data [16, 17, 40, 41]. For scientific problems, one major drawback of the out-of-box machine learning methods is the lack of data. This lack of data has been successfully considered in the geostatistics literature, especially in Bayesian settings [6].

We are interested in improving the ReSL estimates using the limited records of sea level from multiple sources of datasets. Hence, as an extension to our high-dimensional manifold geostatistical model, we provide a novel framework for data fusion of tide-gauge and satellite altimeter to improve the estimates of ReSL.

Intermodel-comparison of Earth system models

Other climate-related variables, such as precipitation, also face challenges that are similar to sea-level changes i.e., non-stationarity, anisotropy, and lack of records [42, 43]. For example, Figure 1.1b shows regional patterns of an annual average of precipitation from the global precipitation climatological project (GPCP) [44, 45].

Furthermore, analysis of a climate-related variable from Earth system models' (ESM) simulations have the potential to complement the observed records when performing statistical inference. One reason to analyze the ESM outputs is that, compared to observed records, simulation outputs from ESM have a lot more data.

Figure 1.3 shows examples of climate model output's precipitation and temperature projections from Community Climate System Model (CCSMv4) [46] around North America. The end of the 21st-century projections shown in Figure 1.3 is for two different greenhouse gas emission (GHG) scenarios.

Climate models, such as CCSMv4, provide simulations of the Earths' past, present, and future states. These fully-coupled climate models are developed by considering earth's important processes of atmosphere, ocean, land surface and sea-ice. Similar to CCSMv4, numerous other international modeling teams are also involved in developing Earth's complex climate models. Collections of these models are assembled by the Coupled Model Inter-comparison Project - 5 (CMIP5) initiative to analyze the effectiveness of various climate models and their projections [47, 48, 49].

One of the scientific questions that CMIP5 is interested in is intermodel comparison [49]. Diagnostic tools, such as intermodel comparison, are crucial for understanding the various causes of disagreements between models. Additionally, the diagnostic tools also aid in evaluating and improving the performance of model projections of climate change.

From a climatological perspective [21, 50], efforts towards categorizing various climate model outputs are made based on their developmental cycles and the different Earth processes that are incorporated into the individual models. Several other work by [24, 51, 52, 53, 54] have developed a multi-variate Bayesian framework to compare

climate models.

To statistically compare climate model outputs it is imperative to consider the data's high-dimensional space. For example, the process sampled at a 1 deg global resolution results in a $360 * 180$-variate analysis. For such high-dimensional spaces, the Euclidean metric is known to be quantitatively ineffective due to it's "curse of dimensionality" [55, 56, 57], while an intrinsic space representation provides a meaningful proximity measure in high-dimensional data sets [58, 59]. Most of the previous approaches in the multi-variate analysis for climate model outputs have not considered the intrinsic space representation of the climate model outputs. Hence, in this thesis, we show a method to represent the climate model output data on an intrinsic high-dimensional space, i.e., a Riemannian manifold, and thereby provide an improved metric to compare climate models.

Emulate future climate projections

Another aspect of the CMIP5 initiative is that of providing an emulator for generating multiple, feasible, future climate scenarios [20, 24]. Future climate scenarios in CMIP5 that depend on GHG emission are also called representative concentration pathways (RCP) [60]. In CMIP5, currently, there are 62 ensemble members from 29 modeling groups that provide some or all of the four different future climate scenarios, i.e., RCP 2.6, 4.5, 6, and 8.5.

By developing a statistical emulator, we can quantify the uncertainty in the existing model outputs, be resourceful in generating the "unseen" future climate states, and, thereby, provide policymakers with a better climate impact analysis [51, 61, 62].

Earth system models (ESM) provide climate model outputs based on various forcing conditions and spatiotemporal scales; but, ESM requires a super-computer that runs for a few months to produce the outputs of interest. In comparison, the statistical emulator that considers the existing ensemble of model outputs provides a differing realization by estimating the ensemble's underlying distribution within a few hours on a laptop.

Realizations of the high-dimensional climate model output require unbiased estimates of the underlying distribution. For an unbiased estimate, it is crucial to consider

inherent dependencies between the ESM models. Dependencies occur due to common code base in the developmental cycle, initialization and parameterization of the geophysical processes, and the usage of common data products [21]. Such inherent dependencies between ESM models and its implication on the climate projection is not comprehensively known [50, 63]. Hence, to circumvent the model dependencies, we provide a clustering framework on the model output's data space, and, thereby, provide unbiased estimates of the ensemble's distribution.

In summary, a typical climate data science cycle, as shown in Figure 1.2, and from a statistical machine learning perspective includes the following steps: 1) collect observations of the natural phenomena, 2) pre-process the datasets, 3) explore the information that serves a particular scientific problem, 4) leverage the exploratory analysis to improve geostatistical models, 5) use improved geostatistical models to provide data-products that explain the reality of the natural phenomena, and 6) finally, communicate the results to decision makers for more effective policy making [20]. This thesis is mainly concerned with the steps three and four as described above.

In the following sections, we provide an outline of the thesis and our contribution of each of the above four scientific goals.

1.4 Thesis Outline

This thesis is organized into seven chapters. In Chapters 2 and 3, we provide a brief introduction to statistical manifold and geostatistics, respectively. In Chapters 4 and 5, we propose non-stationary GP models and further apply to spatial ($m = 2$) and spatio-temporal ($m = 3$) datasets. In Chapters 6, we propose a framework to compare and emulate model outputs from various sources of datasets ($m \geq 50$). Finally, in Chapter 7, we summarize our findings and provide suggestions for future improvements. Summary for each of the chapters is provided below:

In Chapter 2, we give a brief overview of relevant concepts concerning the Riemannian manifold and its application to a statistical manifold for SPD matrices. Specifically, we explain the differential geometry concepts of tangent spaces, connections, and

geodesics. Furthermore, we describe the generalized normal distribution on a statistical manifold, and sketch schemes for estimating the parameters of the distribution. These concepts are applied to improve the geostatistical models.

In Chapter 3, we explain the background of geostatistical model development. The key concepts illustrated in this chapter are the non-stationary covariance functions, geometric anisotropy, multi-variate normal sampling scheme. In the later chapters, a framework for improving these geostatistical concepts is proposed.

In Chapter 4, we propose a framework for estimating the parameters of the spatially-varying non-stationary GP model. The methodology is proposed on spatial and spatiotemporal GP models by using the high-dimensional representation of the geometric anisotropy and a Markov Chain Monte Carlo (MCMC) scheme for finding the GP parameters. We call this model a high-dimensional manifold non-stationary Gaussian Process model (HD-NSGP). Furthermore, we provide simulation studies that emulate the characteristics of a regional climate dataset, such as sea-level changes around North America. For evaluation, we present experiments related to the sea-level change dataset, including tide-gauge observations, satellite altimeter, and Glacial-isostatic adjustment. Finally, the improvement of the proposed model is compared with a climate methodology for estimating the regional sea-level changes [34], a stationary GP model with a Matérn covariance function, and a non-stationary GP method [8].

In Chapter 5, we propose the data-fusion GP model for addressing the scientific problem of inference from multiple sources of datasets. We call this model a data-fused non-stationary GP model (DF-NSGP). Furthermore, we provide an algorithmic procedure for estimating GP parameters for the data-fused, spatially-varying, and non-stationary GP model. The evaluation procedure in this chapter is similar to Chapter 4.

In Chapter 6, we propose a distance function for comparing the ensemble of climate model outputs from CMIP5. Based on the distance function, we then propose a statistical emulator to simulate projections from an existing ensemble of climate model outputs. The dataset chosen for evaluation is that of precipitation changes around North America from the various RCP scenario and projections for 2090. The proposed distance function is compared with the Euclidean distance [64] and a climate genealogy

of CMIP5 [21]. Similarly, the proposed emulator is compared with the vanilla multivariate normal sampling scheme.

In Chapter 7, we conclude by summarizing the findings of the proposed high-dimension manifold geostatistical models given in Chapters 4-6. Furthermore, we provide pointers to the shortcomings and potential improvements in this new approach of developing geostatistical models.

In Appendix A, we provide formulations related to the distance functions for the statistical manifold and spatial processes. In Appendix B, numerical schemes used for the intrinsic statistics and spatial models are sketched out. In Appendix C, notations and acronyms used in this thesis are listed.

Some notes on the notations used in this thesis:

For notations related to geostatistics, we follow Stein [3] and Diggle [1]. For notations related to the statistical manifold, we follow Amari [11]. However, when the notations are conflicting between the two statistical fields, we introduce appropriate statistical-manifold-related notations to maintain consistency with the geostatistics literature.

Where necessary we provide alternative terminologies from the machine learning and climate communities for specific geostatistical definitions. Next, we state some consistent notation style throughout this thesis.

Bold-faced symbols represent vector-valued variables. Calligraphic capital letters are reserved for representing spaces or a set. Subscript indices relate to the variable's input space dimension, e.g., number of samples on a globe, n. Superscript indices refer to the variable's covariate space dimension, e.g., number of samples in the high-dimensional space, m. A hat or a bar on a variable represent estimates from the given samples.

1.5 Thesis Contribution

The overarching contribution of this thesis is in three different areas of research, i.e., geostatistics, climate science, and statistical manifold. The specific contribution of this

thesis is that of representing high-dimensional geometric structures on a statistical manifold, and, thereby, improving the geostatistical models of estimation and simulation. Furthermore, we show the applicability of our proposed models to climate data science problems, such as estimating the regional sea-level changes and simulating future precipitation projections.

Table 1.1 gives a concise breakdown of the original contributions in the above mentioned three fields, its associated references, and pointers to the thesis chapters.

Parts of my contributions in this thesis have been presented in the following papers and are referenced in Table 1.1:

(I) **C. Dalal**, V. Pavlovic, and R. Kopp, *Intrinsic non-stationary covariance function for climate modeling*, ArXiv e-prints,1507.02356, 2015.

(II) **C. Dalal**, V. Pavlovic, and R. Kopp, *Estimating the ocean surface level using the intrinsic non-stationary covariance function*, American Geophysical Union, 2015.

(III) **C. Dalal**, V. Pavlovic, and R. Kopp, *Sea level estimation using the Riemannian manifold and a non-stationary covariance function*, Fifth International Workshop on Climate Informatics, ISBN: 978-0-9973548-0-5, 2015.

(IV) **C. Dalal**, D. Nychka, and C. Tebaldi, *Covariance structure analysis of climate model output*, Sixth International Workshop on Climate Informatics, ISBN: 978-0-9973548-1, 2016.

Table 1.1: Thesis outline and overview of the thesis contributions.

In Chapter	Contributions				
	Dim (m)	Geostatistical Models	Climate Science	Statistical Manifold	In Ref. Above
4	$=$ $2, 3$	• Framework for spatially varying non-stationary Gaussian process (GP) model using a high-dimensional representation of covariate space or geometric anisotropy (HD-NSGP) • Simulation study for non-stationary, spatially-varying, anisotropic, and spatial datasets	Estimating regional sea-level trends without geophysically constraining the dataset	Clustering or sampling symmetric positive definite (SPD) matrices	(I), (II)
5	$=$ $2, 3$	Framework for estimating from multiple sources of non-stationary GPs (DF-NSGP)	Data-fusion of tide-gauge observations with satellite altimeter dataset	A mapping function between two SPD matrix manifold distributions	(III)
6	$\geq$ 50	Framework for comparing ensemble members using the intrinsic space representation of the covariance structure	A metric to compare ensemble members in Climate Model Inter-comparison Project (CMIP5)	Application of Riemannian metric to genealogy	(IV)
6	$\geq$ 50	Framework for a geostatistical emulator	Quickly generate future climate scenarios from existing climate model outputs	Multi-variate normal sampling scheme for high-dimension SPD matrix manifold with dependent samples	(IV)

Chapter 2

Relevant Concepts from High-Dimensional Manifold

2.1 Introduction

In this chapter, we introduce concepts from the statistical manifold literature. For our geostatistical application, we are specifically interested in the multivariate family of distribution on a manifold; hence, we call our framework as the high-dimensional manifold geostatistics.

Intuitively, a manifold is a set of points represented by its coordinate system. When the abstraction of points represents either a probability space or a symmetric positive definite matrix, the manifold is called a statistical manifold. Furthermore, when the probability distribution is from an exponential family and the manifold is equipped with a Riemannian metric, the theory is popularly known as information geometry [11].

For our framework, we restrict ourselves to the exponential families of distribution. One example is the multi-variate normal (MVN) with zero-mean and variance. MVN has a one-to-one correspondence with the symmetric positive definite matrix (SPD) of the same dimension. Note, in this chapter the dimension, m, of a manifold, $\mathcal{S}(m)$, is the number of coordinates required to represent the set of points.

Additionally, we are interested in the inference of complex climate-related objects that are invariant under reparameterization. In other words, we want an estimator for high-dimensional geometric objects that preserve the unbiased and uniformly-minimum variance under the coordinate transformation. The tools developed for these functionally invariant estimators are known as intrinsic statistics [9, 65]. An extrinsic estimator, in comparison, relies on projective spaces such that the manifold is embedded in the ambient Euclidean space. The statistical tools for extrinsic statistics are also commonly

studied under directional statistics [66, 67].

To estimate the underlying distribution of our datasets, which has a small sample size, we are interested in a consistent and well-approximated manifold. Hence, we chose to work with a finite dimensional, geodesically complete, Riemannian manifold [68, 69]. Figure 2.1 illustrates the various concepts of intrinsic statistics on a statistical manifold that we briefly describe in this chapter.

Note, for our notations; we denote a smooth manifold as $\mathcal{S}$. When points on a smooth manifold are a family of probability distributions, we call it the statistical manifold and denote it as $\mathcal{M}$. Furthermore, in this chapter, we relax the bold-faced vector-valued notations when the context is clear.

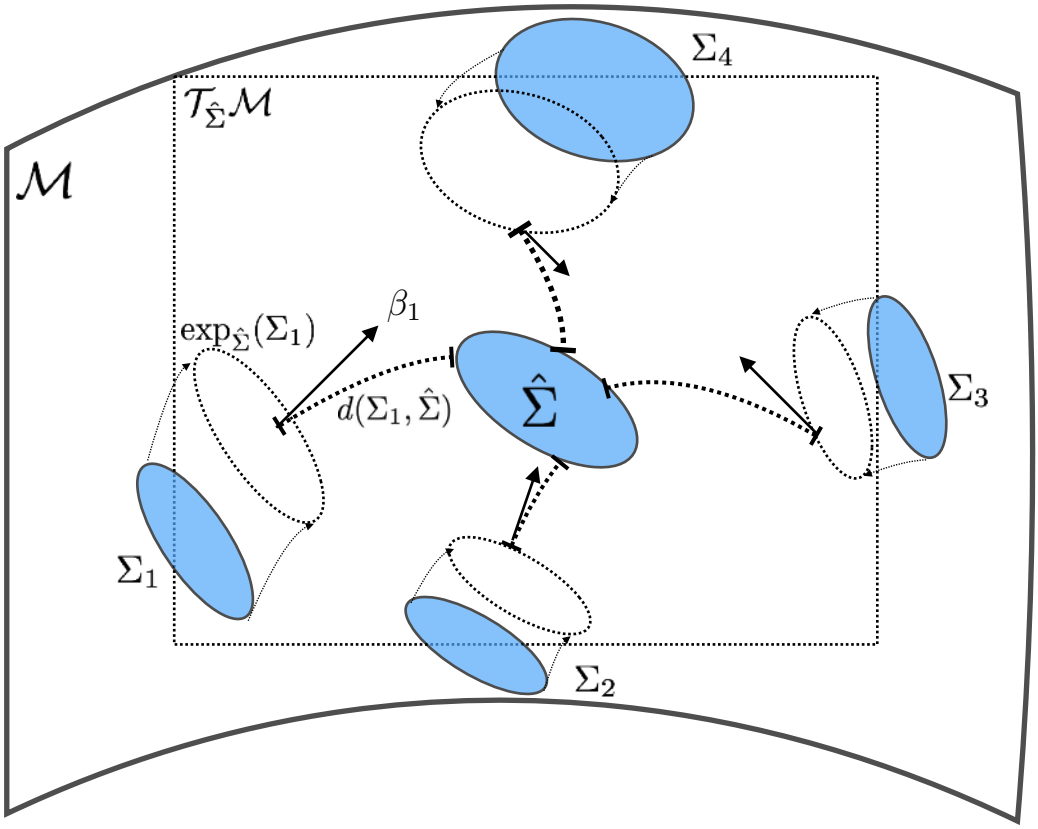


Figure 2.1: A manifold view of high-dimensional geometric objects and their intrinsic statistics. The ellipses depict two-dimensional geometric objects, which are also symmetric positive definite matrices.

The concept of probability distributions on a manifold was first introduced by C.R. Rao in 1947 [12]. Rao made a key observation between differential geometry and the

space of probability distribution by determining a Riemannian metric on a manifold using the Fisher information matrix. Later, Chentsov provided the foundational theorem that the Fisher information metric on statistical models is the only Riemannian metric that is invariant under sufficient statistics [70].

Furthermore, Efron showed the statistical curvature of a one-parameter model and its importance in the asymptotic theory of statistical estimation [71]. Moreover, Dawid introduced the differential geometry concept of connections on the probability spaces [72]. Amari and Nagaoka generalized the information geometry concepts to $\alpha-$connection and dually flat spaces [73, 74, 75, 76].

Recently, the statistical manifold has found a lot of success in medical imaging [9, 10, 77, 78], computer vision [79, 80, 81, 82], and machine learning applications [83, 84, 85, 86]. Amari has also applied the statistical manifold to neural networks using the concepts from multi-terminal information theory and neuro-manifold [11].

For further exposition of various theoretical developments in the statistical manifold literature, we suggest [11], and for differential geometry concepts, we suggest [87].

In the following section, we give an overview of essential differential geometry concepts and their analogous statistical manifold concepts. In Section 2.3, we provide formulations for estimating intrinsic statistics on a statistical manifold, especially, for an MVN family of distribution. Note, the explicit pseudocodes for computing the formulations, that are mentioned in Section 2.3, are outlined in Appendix B.1.

2.2 Riemannian Geometry

2.2.1 Manifold

Definition of a manifold: Let $\mathcal{S}$ be a set. If there exists a set of coordinate systems, $\mathcal{A}$, for $\mathcal{S}$ such that: (i) each element of $\varphi = [\xi^i] = [\xi^1, .., \xi^m]$ in $\mathcal{A}$ and $\psi = [\rho^i] = [\rho^1, .., \rho^m]$ in $\mathcal{A}$ is a one-to-one mapping from $\mathcal{S}$ to some distinct open subset of $\mathbb{R}^m$, and (ii) for all $\varphi, \psi \in \mathcal{A}$, the function $\psi \circ \varphi^{-1}$ is a C^∞ diffeomorphism, then $(\mathcal{S}, \mathcal{A})$ is a m-dimensional C^∞ differential manifold, otherwise known as a smooth manifold. The coordinate transformation $\psi \circ \varphi^{-1} : [\xi^i] \to [\rho^i]$ is a C^∞ diffeomorphism. Therefore,

$\psi \circ \varphi^{-1}$ and its inverse $\varphi \circ \psi^{-1}$ are infinitely many times differentiable.

Definition of a statistical manifold: Let $\mathcal{M}$ be a finite-dimensional manifold, where the points on the manifold are a parametric family of probability distributions, $\{p(x;\xi) : \xi \in \Xi \subset \mathbb{R}^m\}$, and the mapping, $\xi \mapsto p(x;\xi)$, is a diffeomorphism. Then, the statistical models, $p(x;\xi)$, can be described as points on a m-dimensional manifold, $\mathcal{M}$.[1] Here, x is the event space over a finite set, $\mathcal{X}$, of probability distributions, Ξ is the set of coordinate systems, and $\xi = [\xi^1, ..., \xi^m]$ is the real-valued parameters of the probability distribution.

Intuitively, a statistical manifold, $\mathcal{M} = \{p_\xi\}$, describes the continuously-varying parameters, ξ, of the family of probability distributions as the real-valued coordinates of the manifold, $\mathcal{M}$. For example, the multivariate normal distribution with mean, μ, and covariance, Σ, can be described as a m-dimensional manifold, $\mathcal{M}(m, \mathbb{R})$, as follows:

$$\mathcal{M} = \{\xi = [\mu, \Sigma] \quad | \quad \mu \in \mathbb{R}^k, \ \ \Sigma \in \mathbb{R}^{k \times k}, \ \ \forall x \in \mathcal{X}\}, \tag{2.1}$$

where $\mathcal{X} = \mathbb{R}^k$, $m = k + \frac{k(k+1)}{2}$, and $\Sigma \in \text{SPD}$ is a matrix of dimension $\frac{k(k+1)}{2}$.

Note, when $\mu = \mathbf{0}$ points on $\mathcal{M}$ are entirely in the positive real numbers, $\mathbb{R}_+$, and represented by the covariance matrix as $\{\xi = [\Sigma] \in \text{SPD}\,(m = \frac{k(k+1)}{2}) \subset \mathbb{R}_+^m\}$. Figure 2.1 gives a visual depiction of this manifold, $\mathcal{M}$, where the ellipses, Σ, represent the SPD matrices.

One way to represent $\mathcal{M}$ is using functions $\{C, F_i\}$ on $\mathcal{X}$ such that $p(x;\theta) = C(x) + \sum_{i=1}^m \theta^i F_i(x)$. In the literature, this representation is also called a mixture-representation. Another representation that is commonly used for the exponential families of the distribution, such as MVN, Poisson, etc, is the exponential-representation. In this case, $p \mapsto \log p$. Moreover, the statistical manifold of exponential family distributions are also called a flat-manifold, and the m-dimensional manifold, $\mathcal{M} = \{p_\theta\}$, on $\mathcal{X}$ can be written as $p(x;\theta) = \exp\big[C(x) + \theta^i F_i(x) + \psi(\theta)\big]$.[2] Here, $\psi(\theta)$ can be found

[1] In some papers, e.g., in [11], the statistical manifold, $\mathcal{M}$, is also called a statistical model. Additionally, note that the statistical manifold, $\mathcal{M}$, is a subset of family of probability distributions, $\mathcal{P}(\mathcal{X})$, where $\mathcal{P}(\mathcal{X}) \triangleq \{p : \mathcal{X} \mapsto \mathbb{R} \mid p(x) \ (\forall x \in \mathcal{X}), \int p(x)d(x) = 1\}$.

[2] Einstein notation: $\theta^i F_i = \sum_{i=1}^m \theta^i F_i$.

from the normalization $\int p(x;\theta)dx = 1$.[3]

Further generalization to an infinite-dimensional manifold, $\mathcal{S}$, can be carried out by considering the probability distributions over a measurable space ($\mathcal{X}$, $\mathcal{B}$, ν). Here, $\mathcal{X}$ is a set of probability distributions, $\mathcal{B}$ is the Borel field, and ν is the measure for the event space, $\mathcal{X}$. Such spaces are also named as a Fréchet manifold [88].

A useful analogy is analyzing $\mathcal{M}(m,\mathbb{R})$ as the quotient space of $\mathrm{GL}(m)/\mathrm{O}(m)$, i.e., a homogenous space of the Lie group ($\mathrm{GL}(m,\mathbb{R})$). Hence, the statistical manifold can also be analyzed using its associated Lie algebra ($\mathrm{gl}(m,\mathbb{R})$). For analysis, it is further assumed that the support, $supp(P_\xi)$,[4] is constant w.r.t ξ, i.e., for each set of coordinates, ξ, there is one family of probability distributions, $\{p_\xi\}$. Finally, to apply the above theory of the statistical manifold to geostatistical problems, we restrict our analysis to the case of a finite-dimensional manifold that is endowed with a global coordinate system.

2.2.2 Tangent Space

A local coordinate system induces a basis for the tangent space. For example, in Figure 2.1 the neighboring ellipses, Σ_i, are projected onto the tangent space, $\mathcal{T}_{\hat{\Sigma}}\mathcal{M}$, at $\hat{\Sigma}$ using the exponential maps, $\exp_{\hat{\Sigma}}(\Sigma_i)$. In this section, we briefly introduce the tangent space and local projections.

Definition of the tangent space: The m-dimensional tangent space at $p \in \mathcal{S}$, $\mathcal{T}_p\mathcal{S}_m$, is given by the span of m-tangent vectors, $\dot{\gamma}(a) = \dot{\gamma}^i(a)\ (\frac{\partial}{\partial \xi^i})_p$, where $\frac{\partial}{\partial \xi^i}$ is the natural basis of the coordinate system $[\xi^i]$ and $\gamma : I \rightarrow \mathcal{S}$ is a C^∞ curve on the manifold. Here, the curve γ is also a one-to-one function from $I \subset R$ to the manifold $\mathcal{S}$, such that $\gamma^i(t) \triangleq \xi^i(\gamma(t))$ is C^∞ for $t \in I$ and $\dot{\gamma}(a) = (\frac{\mathrm{d}\gamma}{\mathrm{d}t})_p$ for $a \in I$ is the tangent vector at point $p \in \mathcal{S}$. Let c be the collection of all curves passing through the point, p, on the manifold. Then, the tangent space can be written as $\mathcal{T}_p\mathcal{S}_m = \{\ c^i\ (\frac{\partial}{\partial \xi^i})_p \mid [c^i, ..., c^m] \in \mathbb{R}^m\ \}$.

Intuitively, the tangent space $T_p\mathcal{S}$ is derived by locally linearizing the manifold $\mathcal{S}$

---

[3] $\psi(\theta) = \log \int \exp\big[C(x) + \theta^i + F_i(x)\big]dx$.

[4] $supp(p) \triangleq \{x | p(x) > 0\}$.

around p. For example, the tangent plane for a unit sphere can be explained as $\mathcal{T}_p\mathcal{S}_{(m=2)} = \{\dot{\gamma} \in \mathbb{R}^{m+1} \mid \sum_{j=1}^{3} \dot{\gamma}_j \; p_j = 0\}$ or $\{x_0(x - x_0) + y_0(y - y_0) + z_0(z - z_0) = 0 \; : \; x_0, y_0, z_0$ are the coordinates and x, y, z is the point on the sphere$\}$.

Tangent vectors on a smooth manifold can be seen as operators, or directional derivatives, of the function space along the curve $\gamma(t)$. In comparison, for statistical manifolds, the parameterization of the probability model, $p(x;\zeta)$, gives the coordinates of $\mathcal{M}$. Hence, the tangent vectors for the statistical manifold are derivatives of functions in their canonical coordinates, such that $(\partial_i)_\xi = \partial_i p_\xi$ for a mixture (m)-representation and $(\partial_i)_\xi = \partial_i \log p_\xi$ for an exponential (e)-representation.

The tangent space for the m-representation is a linear subspace,[5] such that $\mathcal{T}_p\mathcal{M}_m = \{\mathrm{A} \in \mathbb{R}^{\mathcal{X}} \mid \sum_x \mathrm{A}(x) = 0\}$ and the tangent vectors, $X \in \mathcal{T}_p\mathcal{M}_m$, are elements of the linear subspace, $\sum_x A(x) = 0$. For the e-representation, the tangent space is then described as follows: $\mathcal{T}_p\mathcal{M}_m = \{\mathrm{A} \in \mathbb{R}^{\mathcal{X}} \mid \sum_x p(x)\mathrm{A}(x) = 0\}$. For example, in MVN, i.e., $\{p(x;\theta)\}$, the tangent space, $\mathcal{T}_p\mathcal{M}_m$, is represented by the tangent vectors as:

$$\{\theta^i = \sum_j (\Sigma)^{ij} \; \mu_j \; , \; \theta^{ii} = -\frac{1}{2} \; (\Sigma^{-1})^{ii} \; , \; \theta^{ij} = -(\Sigma^{-1})^{ij} \mid (i < j)\}, \tag{2.2}$$

and the functions on $\mathcal{X}$ are $\{C(x) = 0, F_i(x) = x_i, F_{ij}(x) = x_i x_j : (i \leq j)\}$.

The relation between two tangent spaces, $\mathcal{T}_p\mathcal{S}$ and $\mathcal{T}_{p'}\mathcal{S}$, on a manifold has a structure that is given by the affine connection. Intuitively, the affine connection gives a $1-1$ mapping between $\mathcal{T}_p$ and its neighbors p'. When the manifold is endowed with a Riemannian metric, which we will cover in the next section, the canonical affine connection is also called the Levi-Civita connection, and the curve $\gamma(t)$ joining p and q is called the geodesic curve, i.e., the shortest local distance between the two points on a manifold.

The projection of the neighborhood of points on a manifold, $\mathcal{S}$, to the tangent space, $\mathcal{T}_p\mathcal{S}$, along the geodesic curve, is given by the exponential map.[6] For the curve $\gamma(t)$, if

---

[5]Note, the real-valued function on $\mathcal{X}$ is defined as $\mathbb{R}^{\mathcal{X}} \triangleq \{A | A : \mathcal{X} \mapsto \mathbb{R}\}$ and the affine subspace is defined as $\mathcal{A}_1 \triangleq \{A | \sum_x A(x) = 1\}$.

[6]Exponential map on $\mathcal{S}$ can also be seen as a generalization to the well-known exponential function on $\mathbb{R}$.

$\dot{\gamma}(0) = v$ is the tangent vector at p, such that $\gamma_v(0) = p$, then the exponential map is defined as $\exp_p(v) = \gamma_v(1)$.

Intuitively, the exponential map moves a point on the manifold by a unit distance along the direction of the tangent vectors. For example, if you are standing on the north pole viewing the Earth, then the exponential map of the points on Earth when projected onto the tangent space at the north pole represents all of the longitudinal lines as straight lines. This mapping in cartography is also called the polar azimuthal equidistant projection, and, aptly, is the emblem of the United Nations flag.

Only when the manifold is geodesically complete it's possible to define the exponential map on the entire tangent space. Additionally, the uniqueness of the geodesic curve is derived by assuming the non-existence of cut-locus[7] points on $\mathcal{S}$. For a compact Lie group,[8] the exponential map is also given by $\text{Exp} : \text{gl} \mapsto \text{GL}$, i.e., a map from Lie algebra to Lie group. Conversely, the logarithm map is given by $\text{Log} : \text{GL} \mapsto \text{gl}$.[9]

For $\text{MVN}(\mu, \Sigma)$, with a zero or fixed mean, we can write the family of probability distributions as $p(x, \theta) = \Sigma \in \text{SPD}(m, \mathbb{R}) \leqslant GL(m, \mathbb{R})$, i.e., a space of $m \times m$ real symmetric positive definite matrices.[10] The tangent space is then described by its Lie algebra, $\text{gl}(m, \mathbb{R})$, which is a group of real symmetric matrices, $\text{S}(m, \mathbb{R})$. Moreover, the exponential map coincides with the Riemannian exponential maps and is described as $\text{Exp} : \text{S}(m) \mapsto \text{SPD}(m)$. For the curve, $\Sigma(t)$, where $t \in [a, b] \subset \mathbb{R}$, the formulations for the exponential map are given by $\exp(tX) = \Sigma(a)^{-\frac{1}{2}}\Sigma(t)\Sigma(a)^{-\frac{1}{2}}$ and the tangent vectors are $\dot{\Sigma}(a) = \Sigma(a)^{\frac{1}{2}}X\Sigma(a)^{\frac{1}{2}} \in \text{S}$. Formulations and theorems related to the above expression can be found in [68].

A comprehensive understanding of tangent and projective space requires exposition to tensor field, submanifold, fiber bundle, parallel transport, curvature, and inner product space. For definitions related to differential geometry, we suggest [87, 89], and for

---

[7] Cut-locus are points near p on $\mathcal{S}$ where the geodesic curve is no longer minimizing.

[8] For any compact Lie group, there exist a bi-invariant Riemannian metric.

[9] For notation, we use Exp or $\exp_\Sigma \Sigma$ for the exponential map, and, exp for the natural exponential function.

[10] In group theory, SPD is also denoted as S^+, and is a subgroup of a general linear group, $\text{GL}(m, \mathbb{R})$. Furthermore, the elements of this space, θ, are embedded in $\mathbb{R}^{\frac{m(m+1)}{2}}$.

the analogous techniques related to the statistical manifold, we suggest [11]. We do not provide all of the formal definitions of the manifold so that we can primarily focus on the computational aspect of information geometry. The computational aspect, as elaborated later in Section 2.3, aids us in improving the geostatistical models.

2.2.3 Distance Function

A Riemannian manifold, $(\mathcal{S}, g)$, is a C^∞ differential manifold, $\mathcal{S}$, endowed with a Riemannian metric, g.

Definition of a Riemannian metric: The Riemannian metric, g, is entirely defined by its local inner product on the tangent space, such that $g_p(\cdot,\cdot) : \mathcal{T}_p\mathcal{S} \times T_p\mathcal{S} \to \mathbb{R}$ is symmetric, bi-linear, positive definite, and C^∞ differentiable at $p \in \mathcal{S}$. Additionally, the mapping $g : p \mapsto \langle\cdot,\cdot\rangle_p$ from p to $\mathcal{T}_p\mathcal{S}$ is a tensor field, and the inner product of tangent vectors $X, X' \in T_p\mathcal{S}$ is given by $\langle X, X\rangle_p = g_{ij}(p)X^iX'^j$. Here, $[g_{ij}(p)] = \mathrm{G}(p)$ is a SPD matrix and is also known as a $m \times m$ Gram matrix.

Let $\gamma : [a, b] \to \mathcal{S}$ be the Riemannian curve. Then, its length, or norm, is defined as:

$$\mathrm{L}(\gamma) \triangleq \|\gamma\| = \int_a^b \sqrt{g_{ij}\dot{\gamma}^i\dot{\gamma}'^j}\,\mathrm{dt}, \tag{2.3}$$

and the distance between two points $p, q \in \mathcal{S}$ is $d_g(p, q) = \inf_{\gamma\in\Gamma(p,q)} L(\gamma)$. Here, $\Gamma(x, y)$ is a set of piecewise differentiable curves connecting p and q. The minimum curve, $\inf_{\gamma\in\Gamma(p,q)} L(\gamma)$, on a manifold with an affine connection is called the geodesic distance. For example, on a sphere the geodesic distance is described by the great circle distance.[11]

Following the construction of Rao in [12] and Chentsov in [70], the Gram matrix, $\mathrm{G}(p)$, on a manifold, $\mathcal{S}$, coincides with the Fisher information matrix, $\mathrm{G}(\xi)$, on a statistical manifold, $\mathcal{M}$, such that $g_{ij}(\xi) \triangleq E_\xi[\partial_i l_\xi\ \partial_j l_\xi]$ is the Fisher metric. Here, $l_\xi(x) = \log p(x;\xi)$, and the coordinates, $\{\partial_1 l_\xi, ..., \partial_m l_\xi\}$, is linearly independent.[12] The

[11] Note, geodesic distance is the locally short distance associated to the velocity vector, and not the shortest distance. Intuitively, imagine two points that are connected by an elastic string. Then, the path followed by the tensed string is given by the geodesic curve.

[12] Here, log is the natural logarithm and E_ξ is the expectation w.r.t. $p(x;\xi)$.

Fisher metric is invariant under the coordinate system transformations, uniquely determined, and is the sufficient statistic for the exponential family of probability distributions.

For the MVN distribution, let $\Sigma : \ t \mapsto \Sigma(t) \in \mathcal{M}, \forall t \in [t_1, t_2] \subset \mathbb{R}$ denote a curved segment between two parameterized distributions, $p(.|\Sigma_1)$ and $p(.|\Sigma_2)$. Then, the geodesic distance between the family of multivariate normal distributions, with a common mean vector and different covariance matrices, is given by:

$$D(\Sigma_1, \Sigma_2) = \|\mathrm{Log}_{\Sigma_1}(\Sigma_2)\| = \sqrt{\frac{1}{2}\mathrm{tr}(\log^2(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}))}$$

$$= \sqrt{\frac{1}{2}\sum_{i=1}^{m}\log^2(\lambda_i)}, \tag{2.4}$$

where λ_i denotes the m-eigenvalues of the matrix $(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}) \in S^+$. For visualization, Figure 2.1 depicts the geodesic distance, $d_g(\Sigma_1, \hat{\Sigma})$, between two ellipses, Σ_1 and $\hat{\Sigma}$, on the statistical manifold, $\mathcal{M}$.

Equation 2.4 can be directly derived from Jensen's theorem [90]. Additionally, [68] derives the explicit form, as given in Equation 2.4, of the geodesic distance for the Fisher metric on the manifold $S^+(m, \mathbb{R})$. Note, there are infinitely many connections on a manifold, and, consequently, infinitely many ways to define a Riemannian metric that is associated with the connection. The above definition of the Riemannian metric is associated with the cannonical connection on a manifold $\mathcal{M}$. This canonical connection is also called the Levi-Civita connection, and, Equation 2.4, is also known as the Rao's Riemannian distance. Note, another interesting connection is the $\alpha-$connection, developed by Amari [11], and is out of scope for this thesis.

In Appendix A.2, we discuss various versions of the geodesic distance and some of its properties. Additionally, in Appendix B.1.1 we provide the numerical schemes for computing the Riemannian logarithmic and exponential maps. For further exposition of the distance function, we refer the reader to [68, 77, 90, 91].

2.3 Estimating Intrinsic Statistics

In this section, we define the generalized normal distribution on the statistical manifold for the MVN family of probability distributions, $\{p(\cdot|\Sigma)\}$, and, then, describe schemes for estimating its parameters.[13] Note, as mentioned in Section 2.1, we are interested in an estimator that is invariant under a coordinate transformation. Hence, the tools provided in this section are sometimes categorized in the statistical-manifold literature as the intrinsic statistics.

2.3.1 Generalized Normal Distribution

The generalized normal distribution, $\mathcal{N}(\bar{\Sigma}, \Lambda)$, on the statistical manifold, $S^+(m, \mathbb{R})$, for a mean, $\bar{\Sigma}$, and variance-covariance matrix, Λ, of small variance, $\mathrm{s}^2 = \mathrm{tr}(\Lambda)$, is derived in [9, 68, 77, 91]. The closed form solution of the generalized normal distribution is given by the probability density, $f_G(\cdot)$, of the form:

$$f_G(\Sigma; \bar{\Sigma}, \Lambda_R) = \mathrm{k}_G . \exp \frac{-\boldsymbol{\beta}^T {\Lambda_\mathrm{R}}^{-1} \boldsymbol{\beta}}{2}, \tag{2.5}$$

where the concentration matrix[14] on the Riemannian manifold is given as $\Lambda_\mathrm{R}^{-1} = \Lambda^{-1} - \frac{\mathrm{Ricci}}{3} + O(\mathrm{s}) + \epsilon(\frac{\mathrm{s}}{\mathrm{r_{inj}}})$; the normalizing factor is given as $\mathrm{k}_G = \frac{1+O(\mathrm{s}^3)+\epsilon(\frac{\mathrm{s}}{\mathrm{r_{inj}}})}{\sqrt{(2\pi)^{m(m+1)/2}|\Lambda|}}$; and $\boldsymbol{\beta}$ is the tangent vector from $\bar{\Sigma}$ to Σ.

The curvature tensor, Ricci, is computed at $\bar{\Sigma}$, and, intuitively, provides a measure of the difference between the geometry of the Riemannian metric and the Euclidean metric. The injectivity radius, $\mathrm{r_{inj}}$, at $\bar{\Sigma}$ defines the largest ball around the origin for which the geodesic distance is a locally linearizing function. Finally, the terms $O(\mathrm{s})$ and ϵ are the result of taking the Taylor expansion of the Riemannian measure around the origin such that $\lim_{0+} x^{-\omega}\epsilon(x) = 0 \;\; \forall \omega \in \mathbb{R}^+$. The computation of the Ricci curvature

---

[13] The family of the MVN distribution with a common (or zero) mean, and a variable covariance matrix, Σ, on a statistical manifold, $S^+(m, \mathbb{R})$, can also be defined as a matrix manifold. Given that the point, $p \in S^+(m, \mathbb{R})$, is a symmetric positive definite matrix, $\mathrm{SPD}(m, \mathbb{R})$, and in the quotient space of $\mathrm{GL(m)/O}(m)$.

[14] The inverse of the covariance matrix is also called the concentration matrix, or precision matrix. The first term in ${\Lambda_\mathrm{R}}^{-1}$ is the inverse of the variance-covariance matrix, Λ^{-1}, and the remaining terms, $-\frac{\mathrm{Ricci}}{3} + O(\mathrm{s}) + \epsilon(\frac{\mathrm{s}}{\mathrm{r_{inj}}})$, are the consequence of considering the Riemannian geometry.

tensor for the MVN family is further described in Appendix A.1.

From the derivations in Appendix A.1, it is observed that the Ricci curvature tensor is computationally the most expensive term in Equation 2.5. Hence, an alternative, more straightforward, analytical solution of the above generalized normal distribution can be formulated by assuming that the manifold is locally flat or Euclidean, and geodesically complete. As a consequence, the order of the Ricci curvature is negligible in comparison to the order of variance, $O(s)$, and one can neglect the terms $\frac{\text{Ricci}}{3}$ and r_{inj} in the computation of the generalized normal distribution. The probability density, f_{SG}, for this simplified general normal distribution is then computed as follows:

$$f_{SG}(\Sigma; \bar{\Sigma}, \Lambda) = \text{k}_{SG}. \exp\left(-\frac{1}{2}\text{Log}_{\bar{\Sigma}}(\Sigma)^T \Lambda^{-1} \text{Log}_{\bar{\Sigma}}(\Sigma)\right), \tag{2.6}$$

where the normalizing factor is given as $\text{k}_{SG} = \frac{1}{\sqrt{(2\pi)^{m(m+1)/2}|\Lambda|}}$, and $|\Lambda|$ is the determinant of the variance-covariance matrix, Λ.

Further computational simplification of the general normal distribution can be obtained by using the scalar variance, σ^2, instead of the variance-covariance matrix, Λ. The scalar-variance general normal distribution, $\mathcal{N}(\bar{\Sigma}, \sigma^2)$, with $\sigma > 0$ is then given as the probability density, f_{scG}, as follows:

$$f_{scG}(\Sigma; \bar{\Sigma}, \sigma) = \text{k}_{scG}. \exp\big(\frac{-D^2(\Sigma, \bar{\Sigma})}{2\sigma^2}\big), \tag{2.7}$$

where D is the Rao's Riemannian distance as given in the previous section, i.e., Equation 2.4, and k_{scG} is the normalizing factor. For $m = 2$, the normalizing factor is given as follows:

$$\text{k}_{scG}(\sigma)|_{(m=2)} = \frac{1}{(2\pi)^{3/2}\sigma^2 \ \times \ e^{\sigma^2/4} \ \times \ \text{erf}(\sigma/2)}, \tag{2.8}$$

where erf is the error function. For $m > 2$, the normalizing factor can be computed using a Monte-Carlo simulation for the specific scalar variance, $\sigma \in \mathbb{R}^+$, and, thereby, the pre-computed look-up table can be applied to a particular dataset. Details of the normalizing factor in Equation 2.7 can be found in [92].

The various forms of the generalized normal distribution on a statistical manifold,

as given in Equations 2.5, 2.6, and 2.7, will be used for our geostatisitcal problems of estimation and simulation in Chapters 4, 5, and 6.

2.3.2 Mean and Variance

In this subsection, we describe the estimates of the essential parameters of the generalized normal distribution on a statistical manifold, i.e., the Riemannian mean, $\bar{\Sigma}$, and variance, Λ. Figure 2.1 depicts the empirical Riemannian mean, $\hat{\Sigma}$, and the tangent vector, $\boldsymbol{\beta}_k$, that are used to compute the empirical Riemannian variance, $\hat{\Lambda}$.

Let the empirical Riemannian mean or barycenter, parameterized by $\hat{\Sigma} \in S^+(m, \mathbb{R})$, is estimated from the set of n distributions, $\{p(\cdot|\Sigma_k) : k = 1, \ldots, n\}$, and represented as the normal distributions, $p(\cdot|\hat{\Sigma})$, with a zero mean vector. Then, Fréchet in [88] defines $p(\cdot|\hat{\Sigma})$ as a unique minimizer, $\mathrm{s}_F \in \mathbb{R}^+$, of the Fréchet variance, $\mu_F^2 : S^+(m, \mathbb{R}) \to \mathbb{R}^+$, and is given as:

$$\mu_F^2(\Sigma_1, \ldots, \Sigma_n) = \frac{1}{n} \sum_{k=1}^{n} D^2(\Sigma_k, \hat{\Sigma}) = \mathbb{E}[D^2(\Sigma_k, \hat{\Sigma})],$$
$$\mathrm{s}_F = \mathrm{argmin}_{\,(\hat{\Sigma} \in \mathcal{S}^+)}\, \mu_F^2(\Sigma_1, \ldots, \Sigma_n). \tag{2.9}$$

When $\mu_F^2(\Sigma_1, \ldots, \Sigma_n)$ achieves a local minimum, $\hat{\Sigma} \to \bar{\Sigma}$, it is called the Karcher mean [93]. Karcher showed that for a non-positive sectional curvature, i.e., S^+ manifold, the above mean exists and is unique.

The exponential map and logarithmic map are used in the minimization routine of the gradient descent scheme for numerically finding the empirical Riemannian mean, as outlined in Algorithm 7 in Appendix B.1.2. The numerical scheme for computing the logarithmic map, $\mathrm{Log}_{\hat{\Sigma}}(\Sigma_k)$, is sketched in Algorithm 6. Furthermore, the Riemannian exponential map, $\mathrm{Exp}_{\Sigma_k}(X)$, is inverse of the Riemannian log map, where $\Sigma_k \in S^+(m, \mathbb{R})$ is the initial point on the geodesic flow, $\gamma(t = 0)$, and $X \in \mathrm{SO}(m)$ is the diagonal tangent vector such that $\mathrm{Exp}_{\Sigma_k}(X)$ traverses along the geodesic flow to $\gamma(t = 1)$. Hence, using the similar diagonalization scheme as Algorithm 6, one can compute the exponential map, $\mathrm{Exp}_{\hat{\Sigma}}(\Sigma_k)$.

To find the empirical Riemannian variance or Fréchet variance, $\hat{\Lambda}$, one can associate to each of the n normal distributions, $p(\cdot|\Sigma_k)$, a unique tangent vector, $\boldsymbol{\beta}_k \in \mathrm{S}(m, \mathbb{R})$, at the tangent space, $\mathrm{T}_{\bar{\Sigma}}\mathcal{M}$, such that $\hat{\Sigma}$ is mapped on Σ_k by the exponential map, $\mathrm{Exp}_{\hat{\Sigma}}(\Sigma_k) = \hat{\Sigma}^{1/2} \exp\Big(\hat{\Sigma}^{-1/2}\boldsymbol{\beta}_k\hat{\Sigma}^{-1/2}\Big)\hat{\Sigma}^{1/2}$. The tangent vector, $\boldsymbol{\beta}_k$, is then of the form:

$$\begin{aligned}\boldsymbol{\beta}_k &= -\nabla D^2(\Sigma_k, \hat{\Sigma}) \\ &= -\hat{\Sigma}\log\Big(\Sigma_k^{-1}\hat{\Sigma}\Big),\end{aligned} \tag{2.10}$$

and the variance-covariance matrix is given as:

$$\hat{\Lambda}_{\hat{\Sigma}} = \frac{1}{n-1}\sum_{k=1}^{n}\boldsymbol{\beta}_k\boldsymbol{\beta}_k^T. \tag{2.11}$$

For the scalar version of the empirical Riemannian variance, $\hat{\sigma}$, we can take the maximum likelihood estimates (MLE) of $\log f_{scG}(\Sigma; \bar{\Sigma}, \sigma)$, i.e., Equation 2.7, with respect to the parameters $\bar{\Sigma}$ and σ. The MLE solution for $\hat{\Sigma}$ converges to the empirical Riemannian barycenter of Equation 2.9. Moreover, the MLE for $\hat{\sigma}$ results in the following formulation:

$$\sigma^3 \times \frac{\mathrm{d}}{\mathrm{d}\sigma}\log(k_{scG})^{-1} = \mathbb{E}[D^2(\Sigma_k, \bar{\Sigma})]. \tag{2.12}$$

The above equation can be solved for $\hat{\sigma}_{\hat{\Sigma}}$ using a standard non-linear optimization algorithm, such as Newton's algorithm. Further details of the derivations related to Equation 2.12 can be found in [9, 92].

A crucial practical advantage of using the Riemannian metric, i.e., an affine-invariant metric, and its associated empirical Riemannian mean, is its reduction in the "swelling effects". The "swelling effects" distorts the computation of the high-dimensional tensors, which in our settings is SPD($m \geq 2$). Specifically, the swelling effect is an increase in the determinant of a tensor when computed by averaging or interpolating a set of tensors. The determinant or trace of the tensor is the measure of dispersion in the local diffusion process. Hence, it is essential to select a metric that preserves the determinant of the tensor. In the case of the Euclidean metric, and its associated mean, the swelling

effect is prevalent. For further discussion on the swelling effects, we refer to [10, 94].

In this chapter, we discussed relevant concepts from the statistical Riemannian manifold literature. These concepts are further applied to improve various geostatistical models.

Chapter 3

Relevant Concepts from Geostatistics

3.1 Introduction

The classical geostatistical methodologies that originated in the 1950's from the Fontainebleau school of thought aimed at predicting the yield of the mining operation over a spatially bounded region [13, 14]. However, modern geostatistics has very much integrated with the field of spatial statistics. The three main areas of spatial statistics, that is pointed out in [5], includes continuous spatial variations, discrete spatial variation, and spatial point processes. Furthermore, with the current advent of satellite datasets, *spatial statisticians* have been interested in inference problems in large datasets. Some of the leading inference methodologies for large geospatial datasets include low-rank kriging [95], fixed-rank kriging [96], hierarchical models [6], and, most recently, nearest-neighbor Gaussian process [97].

This thesis mainly concerns with modeling of covariance structures, i.e., the second-order properties of the stochastic process, for the application of continuous spatial variations. For those readers with historic inclinations, extensive development in geostatistics is provided in [1, 5]. Additionally, the exposition of the asymptotic theory in geostatistics can be found in [3].

The classical geostatistical method of Matérn, named Kriging,[1] models a continuous stochastic process, i.e., $\{Z(\cdot)\}$, as:

$$Y(\boldsymbol{s}) \equiv Z(\boldsymbol{s}) + \eta(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{G} \tag{3.1}$$

[1]The geostatitical method of Kriging was named in honor of a mining engineer, D. G. Krige. Specifically, it is called ordinary kriging as the sample mean is estimated using a generalized least squares model. Later extensions, that is universal kriging, used a regression model for the mean estimates.

where $\{Y(\boldsymbol{s})\}$ is the observation process, $\{\eta(\boldsymbol{s})\}$ is the white noise process (or measurement error), and $\{Z(\boldsymbol{s})\}$ is the spatial process which we are interested in modeling. For environmental applications, the spatial processes are classically modeled by their first and second-order moments. The first-order moment being the mean function, i.e., $\mu(\boldsymbol{s}) = E[Z(\boldsymbol{s})] = E[Z(\boldsymbol{s}+\boldsymbol{h})]$, and the second-order moment being the covariance function, i.e., $k(\boldsymbol{s},\boldsymbol{s}') = \text{Cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h})) = \text{Cov}(Z(\boldsymbol{0}), Z(\boldsymbol{h})) = C(\boldsymbol{h})$. Here, $\boldsymbol{h}$ is the inter-point distance vector. This model is also known as a *stationary* process. Furthermore, when $C(h) = C(\|\boldsymbol{h}\|)$ the process is stationary and isotropic, and is also known as homogeneous. Further discussion on the stationary and isotropic processes will be revisited in Sections 3.2.2 and 3.2.3, respectively.

A case of Kriging, specifically, simple kriging, that models a finite-dimensional distribution as a second-order property is called a Gaussian process (GP). GPs are a well known non-parametric regression technique in the geostatistics and machine learning communities and have shown promising results for various real-world applications [98]. The GP models are commonly assumed zero-meaned without loss of generality.

The primary objective of geostatistical modeling, and, in-turn, modeling of the covariance function, is to explain the underlying stochastic process. Foremost step for an explanation, then usually leads to estimating the parameters of the model using the existing record of climate variable, and, next, make inference about the nature of the underlying process of a climate variable, i.e., to find $F = \int_{\mathcal{G}} Z(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$. Here, $\mathcal{G}$ is the spatial domain of the process. The inference stage could also include predictions of the unobserved variables within the spatial domain or could sample a realization of the underlying process for the entire spatial domain. For example, as we show in our methodological development of modeling the covariance function, first we estimate the parameters of the covariance function, $k(\boldsymbol{s},\boldsymbol{s}')$. Then, we make inferences about the climate variables at the unobserved geolocations and regions. The covariance function, intuitively, describes the local, or small-scale, dependence structure of a random variable at various geolocations.

In Chapter 1, we provided an overview of the geostatistical problems describing the focus of this thesis. In this chapter, we delve further into the fundamentals of the

relevant concepts of the geostatistical models for estimation and simulation. Specifically, we describe concepts of the Gaussian process, spatially-varying non-stationary covariance function, geometrical anisotropy, and multivariate normal sampling.

Note, in this chapter, wherever relevant, we point out differences in the usage of terminology between the geostatistics literature and other fields, such as the machine learning literature. One such possible ambiguous terminology is *intrinsic*. Intrinsic, as referred to the statistical manifold, i.e., in Chapter 2, is a class of estimators that is invariant under coordinate transformations. On the other hand, intrinsic random function (IRF) is a class of generalized stochastic processes that are characterized using generalized increments such that its variance is translationally invariant [99, 100].[2] IRF is one of the earliest models in the geostatistics literature for considering non-stationarity of the stochastic process.

3.2 Geostatistical Estimation

In Chapters 4 and 5, we propose an approach for explaining the spatial process of a climate variable, i.e., sea-level change. In this section, we provide the basics that are relevant for the explanation of a spatial process.

3.2.1 Gaussian Process

A stochastic process, $\{Z(\boldsymbol{s}) : \boldsymbol{s} \subset \mathbb{R}^{\dim}\}$, is a *Gaussian process* (GP), such that its finite-dimensional distribution follows a multivariate Gaussian for every set of geolocations, $\boldsymbol{s} = \{\boldsymbol{s_1}, \ldots, \boldsymbol{s_n}\}$, and their associated joint distributions, $\boldsymbol{Z}(\boldsymbol{s}) = \{Z(\boldsymbol{s_1}), \cdots, Z(\boldsymbol{s_n})\}$, i.e., $\boldsymbol{Z}(\boldsymbol{s}) \sim \mathrm{GP}(\boldsymbol{\mu}_Z, \Sigma_Z)$. The standard univariate GP prior model for regression can be expressed as:

$$Y_i \sim \mathcal{N}(Z(\boldsymbol{s_i}), \eta_i) \text{ and } Z(\cdot) \sim \mathrm{GP}(\mu_Z(\cdot), k_Z(\cdot, \cdot, \theta_Y)), \tag{3.2}$$

[2]IRF in classical geostatistics has been used to model the second-order stationarity of the stochastic process. An intuitive idea behind an IRF model is that instead of directly modeling the data, one models a certain linear combination of the data such that the drift is filtered out.

where $\mu_Z(\cdot)$ is the mean function, $k_Z(\cdot,\cdot,\theta_Z)$ is the covariance function, and θ_Z are a set of parameters for the covariance function that is associated with the distribution of $Z(\cdot)$. Furthermore, η_i represents the measurement error, and, in classical geostatistics, the error process is also known as the nugget effect. The standard priors for the measurement error model can be represented as independent, identical, and normally distributed, i.e., $\eta(\cdot) \sim \mathcal{N}(0, \sigma_Z^2)$, where, $\sigma_Z^2 > 0$ is the unknown noise variance (or nugget variance) and $\eta(\cdot)$ is independent of $Z(\cdot)$. Then, the noise covariance matrix is computed as: $K_Y = \sigma_Y^2 \mathbf{I_n}$, where $\mathbf{I_n}$ is the n-dimensional identity matrix.

Notes on the notations: $Y(\cdot)$ denotes functional representation of a process; $\boldsymbol{Y} = (Y(\boldsymbol{s_1}), \cdots, Y(\boldsymbol{s_n}))^T$ represents the vector of observed (or training set) random variables; and, $\boldsymbol{Y}^* = (Y(\boldsymbol{s_1}), \cdots, Y(\boldsymbol{s_{n^*}}))^T$ represents the vector of unobserved (or test set) variables. Similarly, $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$ represents realization of the process over the training and test set, respectively.

From the above GP definition, noisy observed (or training) values, $\boldsymbol{Y}$, and unobserved (or test) values, $\boldsymbol{Y}^*$, can be written in the joint distribution as:

$$\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{Y}^* \end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix} \boldsymbol{\mu}_Z \\ \boldsymbol{\mu}_{Z^*} \end{bmatrix}, \begin{bmatrix} (K_Z + K_Y) & K_{Z^*}^T \\ K_{Z^*} & (K_{Z^{**}} + K_{Y^{**}}) \end{bmatrix} \Big), \tag{3.3}$$

where the covariance matrix K_{Z^*} is computed by applying the covariance function, $k_Z(\cdot,\cdot,\theta_Z)$, between all of the pairs of collections in the observed, $\boldsymbol{s}$, and unobserved geolocations, $\boldsymbol{s}^*$. Similarly, $K_{Z^{**}}$ denotes the computation between $\boldsymbol{s}^*$ and $\boldsymbol{s}^*$.

Gaussian process (GP) regression model have been a useful tool for prediction both in the machine learning community and the spatial statistics community. For the GP model, the predictive distribution is a Gaussian. Furthermore, best linear unbiased prediction (BLUP) for GP is its conditional distribution.[3] Deriving the conditional posterior distribution for $\boldsymbol{Y}^*$ from Equation 3.3, we get a normal with the following

[3]The BLUP, in geostatistics, is generally called Kriging and is derived by minimizing the mean square error (MSE). Furthermore, BLUP in the sense of the best MSE for a GP model coincides with its conditional expectation.

predictive distribution:

$$\boldsymbol{Y}^*|\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{Z^*} + K_{Z^*}^T(K_Z + K_Y)^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_Z), (K_{Z^{**}} + K_{Y^{**}}) - K_{Z^*}^T(K_Z + K_Y)^{-1}K_{Z^*}). \tag{3.4}$$

Given the parameter values, θ_Z, and the noise variance, σ_Y^2, the predictive mean and the variance is given as:

$$\begin{aligned} \mathrm{E}(\boldsymbol{Y}^*|\boldsymbol{Y}, \boldsymbol{\mu}_Z, \theta_Z, \sigma_Y) &= \boldsymbol{\mu}_{Z^*} + K_{Z^*}^T(K_Z + K_Y)^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_Z) \\ Cov(\boldsymbol{Y}^*|\boldsymbol{Y}, \boldsymbol{\mu}_Z, \theta_Z, \sigma_Y) &= (K_{Z^{**}} + K_{Y^{**}}) - K_{Z^*}^T(K_Z + K_Y)^{-1}K_{Z^*}. \end{aligned} \tag{3.5}$$

A typical next step for the methodological development, after settling with the model or hypothesis, is parameter estimation, i.e., finding $\theta = \{\sigma_Y, \theta_Z\}$ from the observed or training datasets. A standard parameter estimation, usually involves some variants of the maximum likelihood-based estimation (MLE) method, i.e., $\theta_{MLE} = \arg\max_\theta \, p(\boldsymbol{Y}|\theta)$. For the Bayesian GP models, the above MLE formulation takes the form of maximizing the posteriors (MAP), i.e., $\theta_{MAP} = \arg\max_\theta \, p(\boldsymbol{Y}|\theta)p(\theta)$.[4] MLE-based methods are desirable for certain classes of models because of their asymptotic properties, such as, consistency and efficiency.

In practice, we have finite samples and complex models that could lead MLE based estimates to over-fitting and analytically-intractable solutions, respectively. Hence, an alternative procedure to use approximate inference methods, such as Markov Chain Monte Carlo (MCMC).

Even though GP is a versatile class of non-parametric model, the key drawback is their computational, $\mathcal{O}(n^3)$, and storage, $\mathcal{O}(n^2)$, complexity. Here, $\mathcal{O}$ is the Big Oh complexity notation, and n is the number of training data points. In recent years, there have been a number of alternative methodologies to circumvent the computation complexity. For example, [97] explores the nearest-neighbor attribute of the spatial process, while [101] exploits the sparsity characteristic of the covariance matrix.

[4]The MLE formulation is equivalent to that of the MAP when the priors on the parameters, $p(\theta)$, are chosen to be uniform.

GPs have been vastly explored in the Bayesian machine learning [98] and spatial statistics [6, 102] literature. The theoretical aspects in terms of fixed-domain asymptotics are provided in [3]. Furthermore, GPs have also been related to other more general non-parametric models, such as Support Vector Machines [103] and smoothing spline models [104]. In this thesis, we do not provide an exhaustive background on the GPs, and, rather, primarily present the GP model for the application of inference using our covariance model.

3.2.2 Spatially-varying Non-stationary Covariance Function

In this section, we discuss relevant basic concepts of a covariance function and then elaborate a specific kind, i.e., spatially-varying non-stationary covariance function. In Chapter 4, other non-stationary models that are relevant to climate applications are further discussed, and a method is proposed to improve the modeling of a spatially-varying non-stationary covariance function.

Covariance Function

A covariance function, $k(\cdot,\cdot)$, on $\mathbb{R}^{\dim} \otimes \mathbb{R}^{\dim}$, of a stochastic process for a pair of geolocations, $\{\boldsymbol{s}, \boldsymbol{s'}\}$, is given as: $k(\boldsymbol{s}, \boldsymbol{s'}) = Cov\{Z(\boldsymbol{s}), Z(\boldsymbol{s'})\} = E((Z(\boldsymbol{s}) - \mu(\boldsymbol{s}))(Z(\boldsymbol{s'}) - \mu(\boldsymbol{s'})))$.[5,6] Moreover, given a set of inputs, $\{\boldsymbol{s_1} \ldots, \boldsymbol{s_n} | \boldsymbol{s_i} \in \mathbb{R}^{\dim}\}$, the entries of the covariance matrix, K, can be computed from from the covariance function, $k(\cdot,\cdot)$, such that $K_{ij} = k(\boldsymbol{s_i}, \boldsymbol{s_j})$. The essential condition on the covariance function, as given in [3], is that $k(\boldsymbol{s}, \boldsymbol{s'}) < \infty$, and for all finite collection of n and $\boldsymbol{a} = \{(a_1, \ldots, a_n)^T | a_i \in \mathbb{R}\}$, it assigns nonnegative values to all quadratic forms, i.e.,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\, a_j\, k(\boldsymbol{s_i}, \boldsymbol{s_j}) = \boldsymbol{a'} K \boldsymbol{a} \geq 0, \quad \forall a_i \in \mathbb{R}. \tag{3.6}$$

[5] A mapping between two inputs $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{I}$ into $\mathbb{R}$ is generally called a kernel function. Here, $\mathcal{I}$ is any arbitrary input space. A symmetric, positive-definite kernel, or the Mercer's kernel, for a pair of inputs is equivalent to the covariance function under certain constraints. Hence, in the machine learning community, the usage of the term covariance function is commonly abused with that of kernel function.

[6] For simplicity, we suppress the inclusion of the probability space notation for a random variable of a stochastic process. That is, instead of $Z(\boldsymbol{s}, \omega)$, where $\omega \in \Omega$ is the sample space, we denote the random variable as $Z(\boldsymbol{s})$.

The above condition on the covariance function further ensures positive definiteness of the correlation function, $\rho(\boldsymbol{s}, \boldsymbol{s'}) := Corr\{Z(\boldsymbol{s}), Z(\boldsymbol{s'})\}$, as is evident from the following formulation: $\rho(\boldsymbol{s}, \boldsymbol{s'}) = \frac{k(\boldsymbol{s},\boldsymbol{s'})}{\sigma(\boldsymbol{s})\sigma(\boldsymbol{s'})}$. Here, $\sigma(\cdot)$, is the positive function, i.e., $\sigma^2(\boldsymbol{s}) = k(\boldsymbol{s}, \boldsymbol{s})$, and $\sigma^2(\boldsymbol{s}) > 0$. Furthermore, the covariance computed between the collection of the random variables is a symmetric, i.e., $k(\boldsymbol{s}, \boldsymbol{s'}) = k(\boldsymbol{s'}, \boldsymbol{s})$.

Intuitively, the covariance function describes how the two random variables change together. Accordingly, the stationary correlation function describes the relationship between two random variables based on their inter-point distance, h. Hence, typically at the origin the stationary correlation function is given as: $\rho(h = 0) = 1$, and it monotonically decays at a distance further away, i.e., $\rho(h \to \infty) \to 0$. Furthermore, when the correlation approaches zero after some finite value of distance, i.e., $\rho(h \geq r) = 0$, then the distance value, $r \in \mathbb{R}^+$, is called the range parameter. Next, we describe a useful form of stationary covariance function which we use in our model development.

Stationary Covariance Function

For modeling real data, the covariance that incorporates two important characteristics of the datasets, i.e., the *closeness* measure, $\rho(\cdot)$, and measurement *noise* model, $\eta(\cdot)$, can be written as:

$$k(\boldsymbol{s_i}, \boldsymbol{s_j}) = \sigma_Z^2 \, \rho(q(\boldsymbol{s_i}, \boldsymbol{s_j})) + \sigma_Y^2 \mathrm{I}, \tag{3.7}$$

where $\sigma_Z^2 > 0$ is the signal variance, $\sigma_Y^2 > 0$ is the noise variance, and I is the indicator function. In the above formulation, we have assumed a constant variance leading to $\sigma^2 = \sigma(\boldsymbol{s})\sigma(\boldsymbol{s'})$. In geostatistics, the noise variance, σ_Y^2, is also known as the nugget variance, and the variances, $\sigma_Z^2 + \sigma_Y^2$, is also known as the sill variance.

The key component in the correlation function that measures *closeness* is the scaled squared separation length (henceforth, called the dissimilarity function), $q_{ij} \triangleq q(\boldsymbol{s_i}, \boldsymbol{s_j})$, and is given as:

$$q(\boldsymbol{s_i}, \boldsymbol{s_j}) = (\boldsymbol{s_i} - \boldsymbol{s_j})^T \Sigma_Z^{-1} (\boldsymbol{s_i} - \boldsymbol{s_j}), \tag{3.8}$$

where Σ_Z is a positive definite matrix, and, henceforth, is called the scale parameter.[7]

When the process is assumed to have a constant scale in every direction of the input space axes, the above formulation can be written as: $q_{ij} = \frac{d^2(\boldsymbol{s_i},\boldsymbol{s_j})}{l^2}$. Such a process is also called geometrical isotropic. Here, $l \in \mathbb{R}^+$ is the scalar scale parameter, and $d_e(\cdot,\cdot)$ is the distance function between a pair of geolocations.[8,9] On the other hand, when the process is allowed to have directional effects with respect to the scale parameter, i.e., geometrical anisotropy, the scale parameter is a matrix, i.e., $\Sigma_Z \in \mathbb{R}^{\dim\times\dim}$. Furthermore, when the feature, or covariate space, of s is considered, i.e., $\boldsymbol{s} \to \boldsymbol{x}(\boldsymbol{s})$, the scale matrix dimensions are given as: $\Sigma_Z \in \mathbb{R}^{m\times m}$.

For spatial statistics, including geostatistical models, an important class of covariance function is the Matérn function due to their flexibility in capturing the varying smoothness of the underlying distribution. The Matérn form of the stationary correlation function, M_ν, can be constructed as follows:

$$\rho(\nu, q_{ij}) = \frac{1}{\Gamma(\nu)2^{\nu-1}}(\sqrt{2\nu q_{ij}})^{\nu} K_\nu(\sqrt{2\nu q_{ij}}), \tag{3.9}$$

where ν is the smoothness parameter that controls the differentiability of the function, K_ν is the modified Bessel function of the second kind, and $\Gamma(\nu)$ is the gamma function. Note, when $\nu \to \infty$ the Matérn function approaches the squared exponential correlation function, i.e., $\rho(q_{ij}) = \exp\left(\frac{-q_{ij}}{2}\right)$.[10] Similarly, when $\nu = 0.5$ the Matérn function approaches the exponential function, i.e., $\rho(q_{ij}) = \exp\left(-\sqrt{q_{ij}}\right)$.

In classical geostatistics, the second-order dependence in a stochastic process has also been modeled using the Variogram model. The theoretical variogram, $V(\cdot,\cdot)$, is

---

[7]Interestingly, the above form of the range function, $q(\boldsymbol{s},\boldsymbol{s'})$, is equivalent to that of the Mahalanobis distance, or the generalized squared inter-point distance between the observation vectors, i.e., $d^2(\boldsymbol{s},\boldsymbol{s'}) = (\boldsymbol{s}-\boldsymbol{s'})^T\Sigma^{-1}(\boldsymbol{s}-\boldsymbol{s'})$. Here, $d(\cdot,\cdot)$ is the dissimilarity measure, and Σ is the positive definite covariance matrix.

[8]In geostatistical models, the role of the scale parameter, l or Σ_Z, is similar to that of the range parameter, r. Additionally, for the exponential correlation function the scalar scale parameter, l, can be trivially seen as equivalent to the range parameter, r.

[9]In the machine learning community, the scalar scale parameter, l, is also called the characteristic length scale.

[10]In the machine learning community, a squared exponential correlation function is also popularly called as the radial basis function.

given as:

$$V(\boldsymbol{s}, \boldsymbol{s}') = \frac{1}{2} Var(Z(\boldsymbol{s}) - Z(\boldsymbol{s}')). \tag{3.10}$$

For practical purposes, the empirical variogram, $V_Y(\boldsymbol{h}) = \frac{1}{2}E[(Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h}))^2]$, is commonly used. However, for inference, several authors, including [1, 3], suggested avoiding the empirical variogram model, and, instead, using the Matérn model. Albeit, due to the simplicity of the empirical variogram model, it is regularly used for exploratory purposes.

When the covariance is only a function of the vector difference between the geolocations, i.e., $k(\boldsymbol{s}, \boldsymbol{s}') = k(\boldsymbol{s} - \boldsymbol{s}')$, and the mean of the process is constant, $\mu(\boldsymbol{s}) = \mu$, the process is called weakly stationary.[11] On the other hand, any stochastic process that does not follow the above two conditions is called a non-stationary process. In spatial statistics, non-stationarity is commonly introduced by varying the mean structure, or covariance structure, or by using the intrinsic random function (IRF).

Non-stationary Covariance Function

Non-stationarity could simply be introduced by varying the mean based on its location, i.e., $\mu(\boldsymbol{s})$. Such mean functions are called spatial trends, and can be easily modeled using either the polynomial regression or spatially referenced covariates [1]. Alternatively, IRF models non-stationarity in the process by using stationary increments. A straightforward example is that of a random walk model. IRF models are not widely used in environmental applications, especially for spatial predictions, as they fail to capture the global behavior of the process.

In environmental applications, a crucial non-stationarity is that of the covariance structure [5]. The general philosophy underlying most of the modeling approaches for the non-stationary covariance structure is that *locally* the correlation structure is approximately stationary and anisotropic. Such covariance structures are characterized

[11] A weakly stationary process implies shift invariance of its first and second-order moments. In comparison, a strictly stationary process implies shift invariance of the finite-dimensional distribution. In general, a strict stationarity implies a weakly stationary process, but the converse is not true. However, for the stationary Gaussian stochastic process the converse is also true.

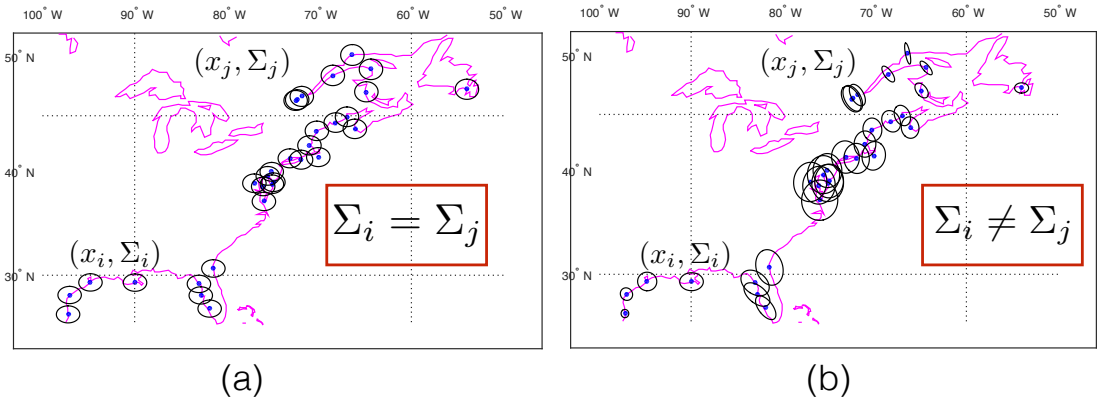


Figure 3.1: Example of geometrical anisotropy (i.e., ellipses) along Northeast America's tide-gauge sites. The shape of the ellipses are (a) constant for the stationary covariance structure and (b) a function of geolocations for the spatially-varying non-stationary covariance structure.

as locally-stationary, anisotropic, and spatially-varying.

For an illustration of the spatially-varying *non-stationary* covariance structure see Figure 3.1b. Here, the shape of the ellipses at various tide-gauge locations around Northeast America is varying. Note, the ellipses are representing the spatial anisotropy locally, i.e., Σ_Z in Equation 3.8. In comparison, the anisotropic *stationary* covariance structure is illustrated in Figure 3.1a, where the ellipses are constant over the spatial region.

In the literature of spatial statistics, predominantly the non-stationary models for the covariance structure can be categorized into smoothing kernel methods, basis function models, deformation models, and process convolution models. For the basics on the aforementioned models see [5]. Further discussion of the literature of non-stationary models is revisited in Section 4.1. In this section, we provide the background for the process convolution models as we propose to improve these class of models in Section 4.3.1.

The process convolution approach, as introduced by [7], models the process, $Z(\boldsymbol{s})$, by convolving the kernel function, $k_{\boldsymbol{s}}(\cdot)$, centered at the locations $\boldsymbol{s}$ with the white noise process, $w(\cdot)$, i.e., $Z(\boldsymbol{s}) = \int_{\mathcal{G}} k_{\boldsymbol{s}}(\boldsymbol{u})w(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$. Here, the locations are $\boldsymbol{s}, \boldsymbol{u} \in \mathcal{G} \subset \mathbb{R}^{\dim}$. By allowing the kernels to vary spatially, the resulting non-stationary covariance

function is then given as:

$$k(\boldsymbol{s}, \boldsymbol{s}') = \int_{\mathcal{G}} k_{\boldsymbol{s}}(\boldsymbol{u}) k_{\boldsymbol{s}'}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}. \tag{3.11}$$

For the closed-form solution of the above convolution model, [7] used Guassian kernels that vary smoothly in the input space. Further improvement by [8] modeled the kernels as spatially-varying parameters of the Matérn covariance function.

The closed-form solution of the spatially-varying non-stationary Matérn correlation function, $\rho^{NS}(\cdot, \cdot)$, as given by [8], is:

$$\rho^{NS}(\boldsymbol{s_i}, \boldsymbol{s_j}) = 2^{\frac{\dim}{2}} |\Sigma_i|^{\frac{1}{4}} |\Sigma_j|^{\frac{1}{4}} \left|\frac{\Sigma_i + \Sigma_j}{2}\right|^{-\frac{1}{2}} \rho^{S}(\nu, q_{ij}^{NS}), \tag{3.12}$$

where $\Sigma_i \triangleq \Sigma(\boldsymbol{s_i})$ is the spatially-varying geometrical anisotropy, and $\rho^S(\cdot, \cdot)$ is the stationary Matérn correlation function as given in Equation 3.9. The dissimilarity function, q_{ij}^{NS}, is given as:

$$q_{ij}^{NS} = (\boldsymbol{s_i} - \boldsymbol{s_j})^T \Big(\frac{\Sigma_i + \Sigma_j}{2}\Big)^{-1} (\boldsymbol{s_i} - \boldsymbol{s_j}), \tag{3.13}$$

where the scale matrix, $\Sigma(\boldsymbol{s})$, is positive definite, i.e., $\Sigma(\boldsymbol{s}) : \mathbb{R}^{\dim} \rightarrow \mathrm{SPD}(\dim, \mathbb{R})$, and, in practice, preferred to be smoothly varying. Furthermore, by the property of SPD matrices, the convolved scale matrix, or the geometrical anisotropy, $\Sigma_{ij} \triangleq \Sigma(\boldsymbol{s_i}, \boldsymbol{s_j}) = \frac{\Sigma_i + \Sigma_j}{2}$, is also positive definite. Note, the correlation function, $\rho^{NS}(\cdot, \cdot)$, leads to a valid non-stationary covariance function, and the proof can be found in [8]. Intuitively, Equation 3.12 indicates that the non-stationary correlation structure between $\boldsymbol{s}$ and $\boldsymbol{s}'$ is captured by the convolution (or the arithmetic average) of the local kernel structure (or the geometrical anisotropy).

Moreover, extension of Equation 3.12 is provided by [18], where in addition to the spatially-varying geometrical anisotropy, $\Sigma(\boldsymbol{s})$, the smoothness parameter of the Matérn correlation function is also spatially-varying, $\nu(\boldsymbol{s})$.[12] The non-stationary covariance

[12]The formulation of Equation 3.14 is no longer a process convolution model, but, it is still a valid non-stationary covariance function.

function, $k^{NS}(\cdot,\cdot)$, as given in [18], can be constructed from the following form:

$$k^{NS}(\boldsymbol{s_i},\boldsymbol{s_j}) = c(\boldsymbol{s_i})c(\boldsymbol{s_j})|\Sigma_{ij}|^{-\frac{1}{2}}\ \rho^S\left((\tfrac{\nu(\boldsymbol{s_i})+\nu(\boldsymbol{s_j})}{2}), q_{ij}^{NS}\right), \tag{3.14}$$

where $c(\boldsymbol{s_i})$ and $c(\boldsymbol{s_j})$ are functions on $\mathbb{R}^{\dim}$, such that $c(\boldsymbol{s_i}) = f(\boldsymbol{s_i}, |\Sigma_i|^{1/4})$ and $c(\boldsymbol{s_j}) = f(\boldsymbol{s_j}, |\Sigma_j|^{1/4})$. This equation reduces to Equation 3.12 when a constant smoothness and signal variance is assumed, i.e., $\nu(\boldsymbol{s}) = \nu$ and $c(\boldsymbol{s}) = \sigma|\Sigma(\boldsymbol{s})|^{1/4}$.

Process on the Sphere

For environmental applications, the curvature of the Earth should be considered while modeling processes that are located far apart in the input space [105]. Such processes are assumed to lie on a sphere, $\mathcal{S}^2 \subset \mathbb{R}^3$. For processes on a sphere, the inter-point distance between geolocations are computed using the great circle distance, or, commonly, the angular distance, d_{gcd}.[13] See Appendix A.3 for the computational details of the distance function on a sphere.

For the stationary isotropic correlation, the dissimilarity function considering the great circle distance, i.e., Equation 3.8, can be written as: $q_{ij}^t = \frac{d_{gcd}^2(\boldsymbol{s_i},\boldsymbol{s_j})}{l^2}$. Furthermore, for the case of a geometrical anisotropy correlation function, the input space of geolocations can be transformed, $\boldsymbol{s} \to \boldsymbol{s^t}$, such that, $\boldsymbol{s^t} = L\boldsymbol{s}$ and, $LL^T = \Sigma^{-1}$. Then, the transformed dissimilarity function can be written as:[14]

$$q_{ij}^t = d_{gcd}^2(\boldsymbol{s_i^t}, \boldsymbol{s_j^t}). \tag{3.15}$$

For the case of a spatially-varying non-stationary geometrical anisotropy correlation function, the dissimilarity function in Equation 3.13 is a function of the convolved scale matrix, Σ_{ij}, and depends on both of the input space locations, $\boldsymbol{s_i}$ and $\boldsymbol{s_j}$. Hence, the

[13]Great circle distance can be given as: d_{gcd} = rad Δ central angle, where the radius, *rad*, of the earth is assumed constant. Hence, the great circle distance between geolocations, in terms of latitude and longitude, is typically computed with the angular distance, i.e., Δ central angle, in the unit of radians.

[14]In [19], a deformation model, similar to the transformed stationary anisotropic correlation function, is devised for the non-stationary anisotropic covariance function.

transformation scheme of Equation 3.15 cannot be directly applied for q_{ij}^{NS} as it renders to a semi-metric dissimilarity function and, potentially, a discontinuous scale matrix process on a sphere.[15] In that case, [8] suggests using the Lambert azimuthal equidistant Euclidean projection of the geolocations on a sphere for modeling a smooth scale matrix process. The scale matrices are then projected back to the sphere for modeling the convolved correlation function. As pointed out in [8], such procedures do not have an analytical form, and the numerical integration is not straightforward. Hence, in practice, an Euclidean distance approximation of the spherical distance suffices for regional datasets, such as the North American tide gauge sites.

Smoothness Properties

Moreover, the choice of the covariance function for the Gaussian process and its local behavior affects the critical smoothness properties of the underlying physical or stochastic process. Specifically, the two critical properties of a process are its mean square and sample path continuity and differentiability. The basic theory of the smoothness properties of a process can be found in [106]. For a standard zero-mean Gaussian process, the various covariance functions given above are mean square and sample path continuous. For the stationary Matérn correlation function of Equation 3.9, the process is $2M$-th and M-th order mean square and sample path differentiable, respectively, if the smoothness parameter is $\nu > 2M$.[16] Further discussion of the properties of the stationary Matérn correlation function can be found in [3].

For the spatially-varying non-stationary correlation function of Equation 3.12, the mean square and sample path is guaranteed M-th order differentiable if the smoothness parameter is $\nu > 2M$ and the anisotropic scale matrix process, $\Sigma(\boldsymbol{s})$, has $(M+1)$-th order differentiability. Hence, in practice a smoothly varying geometrical anisotropy is preferred for the spatially-varying non-stationary correlation function. Intuitively, the non-stationary covariance function modeled through a smooth anisotropic process is

[15] A semi-metric violates the triangle inequality condition of the distance function

[16] The covariance function form of Equation 3.7 is assumed to have constant variances, $\sigma_Y^2 = \sigma_Y^2(\boldsymbol{s})$ and $\sigma_Z^2 = \sigma_Z^2(\boldsymbol{s})$; hence, they do not affect the smoothness properties of the process.

locally stationary; hence, one can see the similarity between the smoothness properties of the above given stationary and non-stationary correlation functions. The proofs of the aforementioned smoothness properties are given in [8]. In the next section, we provide methods from the literature for modeling a smoothly-varying geometrical anisotropy, or the scale matrix.

3.2.3 Geometrical Anisotropy

Anisotropy mean that the properties of a process are directional dependent. An important anisotropy that can be introduced through the covariance structure is the geometrical anisotropy.[17] *Geometrical anisotropy*, intuitively, explains the differential rate of correlation decay between the geolocations relative to their orientation [1]. For the stationary case, as mentioned in Equation 3.8, geometrical anisotropy through the scale matrix, in $\dim = 2$, can be decomposed as:

$$\Sigma_Z = \Gamma \mathrm{D} \Gamma^T = \begin{bmatrix} \cos(\psi_A) & -\sin(\psi_A) \\ \sin(\psi_A) & \cos(\psi_A) \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \cos(\psi_A) & \sin(\psi_A) \\ -\sin(\psi_A) & \cos(\psi_A) \end{bmatrix}. \tag{3.16}$$

Here, Γ is the eigenvector matrix, or the rotation matrix, that can be expressed using the anisotropy angle, ψ_A, and D is the diagonal matrix with positive eigenvalues, $\{\lambda_1, \lambda_2 \mid \lambda_i \in \mathbb{R}^+\}$, that can also be expressed in terms of the anisotropy ratio, $\psi_R = \frac{\lambda_1}{\lambda_2} \geq 1$. Note, this decomposition of a two-dimensional SPD matrix, Σ_Z, can be represented as an ellipse, as depicted in Figure 3.1.

For the case of a non-stationary covariance structure, the geometrical anisotropy is modeled such that it varies for every pair of geolocation, Σ_{ij}. For example, the geolocations that are associated with a large anisotropy ratio has a slow rate of correlation decay along the direction of the anisotropy angle.

The geometrical anisotropic process, introduced by [7] as the kernel convolution process, uses Gaussian kernels for each geolocation, $\boldsymbol{s_i} \in \mathbb{R}^{(\dim=2)}$ to model the scale

[17] Similar to the geometrical anisotropy in the correlation function, anisotropy can also be introduced in the sill and nugget variances. In this thesis, we focus on the most important anisotropic parameter, i.e., geometrical anisotropy. For discussion on other anisotrppic properties of a process, such as, zonal and nugget anisotropy, see [1, 102].

matrix.[18] They model the smoothly-varying Gaussian kernels using the parameterization of an ellipse with a constant area. Such parameterizations are not straightforward to be scalable for high-dimensional spaces. Furthermore, a constant area of the ellipse and Gaussian kernels can cause over-fitting and over-smoothing of the sample paths, respectively.

An extension provided by [8] models geometrical anisotropy as the average of the scale matrices that are associated with the pair of geolocations, i.e., $\Sigma_{ij} = \frac{\Sigma(\boldsymbol{s_i})+\Sigma(\boldsymbol{s_j})}{2}$, as given in Equation 3.13. They evaluate the scale matrix for each of the geolocations using eigendecomposition, and, in-turn, model each of the parameters of the decomposed matrix using a stationary anisotropic process. The eigendecomposition form for the two dimensional input space in [8] is as follows:

$$\Gamma(s) = \begin{bmatrix} \frac{u}{l_{uv}} & \frac{-v}{l_{uv}} \\ \frac{v}{l_{uv}} & \frac{u}{l_{uv}} \end{bmatrix}, \ \mathrm{D}(s) = \begin{bmatrix} \log(\lambda_1) & 0 \\ 0 & \log(\lambda_2) \end{bmatrix}, \ l_{uv} = \sqrt{u^2 + v^2}, \tag{3.17}$$

where the parameters of the scale matrix, $\theta_{\Sigma_Z} \triangleq \{\theta^i_{\Sigma_Z}\}_{\boldsymbol{s}} = \{u(\boldsymbol{s}), v(\boldsymbol{s}), \lambda_1(\boldsymbol{s}), \lambda_2(\boldsymbol{s})\}$, are modeled using an independent stationary anisotropic Gaussian Process. In other words, each parameter, $\theta^i_{\Sigma_Z}$, of the above eigendecomposition is modeled as:

$$\theta^i_{\Sigma_Z}(\cdot) \sim \mathrm{GP}(\mu^i_{\Sigma_Z}, k^i_{\Sigma_Z}(\cdot, \cdot, \theta_{\theta^i_{\Sigma_Z}})), \tag{3.18}$$

where $\theta_{\theta^i_{\Sigma_Z}} \triangleq \{\theta_{\theta^i_{\Sigma_Z}}\} = \{\sigma^2_{\theta^i_{\Sigma_Z}}, \ \Sigma_{\theta^i_{\Sigma_Z}}\}$ are the hyperparameters, or free parameters, for the covariance, $k^i_{\Sigma_Z}$, of the GP model that is associated to the scale matrix parameter of $\theta^i_{\Sigma_Z}$. To ensure a consistent sample path differentiability along the model hierarchy, a stationary squared exponential correlation function, i.e., infinitely differentiable, is used by [8] for $k^i_{\Sigma_Z}(\cdot)$. Furthermore, depending on the application, for modeling the hyperparameters, $\theta_{\theta^i_{\Sigma_Z}}$, a uniform prior or conjugate priors should suffice. Note, for the notations in the above parameters, the superscript denotes an element of the parameter

[18] In the literature of geostatistics, e.g., in [7, 8, 107], the terminology of kernel matrix is used for the spatially-varying geometrical anisotropy. Instead of using the term kernel matrix, we use scale matrix, so as not to be confused with the widely, and sometimes loosely used term of kernel in the machine learning community.

space at each layer of the model hierarchy, and the subscript denotes the associated process of a parameter that is a level above.

The above model of the geometrical anisotropy, as given by [8], has three layers of parameters in the model hierarchy. The first layer closest to the data is the parameters of variances and smoothness, i.e., $\theta_{\text{level}=1} = \{\sigma_Y^2, \sigma_Z^2, \nu\}$; the second layer is the scale matrix process, i.e., $\theta_{\text{level}=2} = \{\theta^i_{\Sigma_Z}\}_{\boldsymbol{s}\in\mathbb{R}^{\text{dim}}}$, where $i = 1, \ldots, \Big(\text{dim} + \frac{\text{dim}(\text{dim}-1)}{2} + (\text{dim} - 1)\Big)$; and, finally, the last layer of the hyperparameters are $\theta_{\text{level}=3} = \{\theta_{\theta^i_{\Sigma_Z}}, \mu^i_{\Sigma_Z}\}$. Hence, the number of parameters to be estimated, $|\theta_{\text{est}}|$, discounting the parameters of the priors for $\theta_{\text{level}=1}$ and $\theta_{\text{level}=3}$, can be written as follows:

$$|\theta_{\text{est}}| = \Big(3\Big) + \Big(2 \cdot \text{dim} - 1 + \frac{\text{dim}(\text{dim}-1)}{2}\Big) \cdot \Big(\text{dim} + \frac{\text{dim}(\text{dim}-1)}{2} + 1\Big). \quad (3.19)$$

Note, the above cardinality only refers to the parameters that need to be estimated. More specifically, the total number of the parameters in the model also includes the number of scale matrix, n, computed at each of the observations. Hence, the dimension for the parameter space, discounting the priors, is $|\theta_{\text{total}}| = (|\theta_{\text{est}}| - 3) \cdot n + 3$. Furthermore, the computations complexity for the first level is $\mathcal{O}(n^3)$, and for the second level, it is $\mathcal{O}(\text{dim}^2 \cdot n^3)$.

The non-stationary model described in [8], allows for a greater degree of flexibility in modeling the smoothly-varying scale matrix. Yet, one major drawback of the above model, especially, for applications of real datasets, is the estimation of the many parameters of various processes in the model hierarchy. This drawback could lead to overfitting and mixing issues, as pointed out by [107, 108]. As can be seen from Equation 3.19, the order of the parameter space is $(\text{dim}^4 \cdot n)$, and when dim = 2 the cardinality of the parameters to be estimated is $|\theta_{\text{est}}| = 19$. Furthermore, in [8] the extension of the above model to $\boldsymbol{s} \in \mathbb{R}^{(\text{dim}=3)}$ is prescribed, but not recommended, due to its infeasibility in estimating a high-dimensional parameter space.

A more recent approach in [107] models the spatially-varying geometrical anisotropy using a covariate based regression, i.e., $\Sigma(\boldsymbol{s}) = \Psi + \Gamma\boldsymbol{x}(\boldsymbol{s})\boldsymbol{x}(\boldsymbol{s})^T\Gamma^T$. Here, Ψ represents the error covariance matrix, and Γ refers to the variability matrix in the covariate

space, m. Here, the total number of parameters in modeling the geometrical anisotropy is reduced from the number of observations, n, to that of the dimension of the covariate space, m. Even though the restriction of the parameter space as a function of the covariate space by [107] is useful, it is unclear how many attributes of the covariate space one requires to explain the local stationarity in the input space of the dataset. Moreover, parameter estimation of the various levels in the model hierarchy remains a challenging task, especially in a high-dimensional covariate space. In Chapter 4, we propose a non-stationary covariance function to circumvent some of the issues with the spatially-varying geometric anisotropy in a high-dimensional covariate space.

3.3 Geostatical Simulation

In Chapter 6, we propose a method to compare Earth system models and simulate future projections of climate-related variables. To capture the variability of future climate projections at locations around the globe and across ensemble members, one requires a multivariate model framework. Hence, we employed the basic model of multivariate normal (MVN) sampling to show the validity of our approach in drawing samples from a MVN distribution on a manifold of covariance structure. In this section, we briefly describe relevant concepts and establish the notations pertaining to Chapter 6.

If $\boldsymbol{Z(s)} = \{Z_1(\boldsymbol{s}), \ldots, Z_n(\boldsymbol{s})\}$ follows the standard multivariate normal, MVN$(\boldsymbol{\mu}, \Sigma)$, and assuming there exists an affine transformation, the multivariate samples, $\tilde{\mathbf{z}}$, can be drawn as follows:

$$\tilde{\mathbf{z}} = \boldsymbol{\mu} + \Sigma(\theta)^{\frac{1}{2}}\boldsymbol{\epsilon}, \tag{3.20}$$

where $\boldsymbol{\mu}$ is the multivariate mean, Σ is the multivariate covariance matrix, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the standard normal random noise vector. Assuming independence, the parameters of the distribution can be estimated as follows:

$$\hat{\boldsymbol{\mu}} = \frac{1}{\mathrm{n}}\sum_{i=1}^{\mathrm{n}} \mathbf{z_i}\,, \quad \hat{\Sigma}(\theta) = \sigma^2_{\boldsymbol{Z}} H(\Sigma_{\boldsymbol{Z}}) + \sigma^2_{\boldsymbol{Y}}\mathbf{I}. \tag{3.21}$$

In this preliminary study, the correlation matrix, H, is computed from a stationary

multivariate correlation function. The parameters, $\theta = \{\sigma_{\boldsymbol{Y}}, \sigma_{\boldsymbol{Z}}, \Sigma_{\boldsymbol{Z}}\}$, of the covariance matrix, i.e., noise variance, signal variance, and scale matrix, are estimated by maximizing the likelihood function (MLE) and using standard optimization techniques. The likelihood function for the standard MVN is of the form:

$$\mathcal{L}(\theta|\mathbf{z_1}, \ldots, \mathbf{z_n}) \propto |\Sigma(\theta)|^{-\frac{\mathrm{n}}{2}} \prod_{i=1}^{\mathrm{n}} \exp\left(-\frac{1}{2}(\mathbf{z_i} - \hat{\boldsymbol{\mu}})^{\mathsf{T}} \Sigma(\theta)^{-1} (\mathbf{z_i} - \hat{\boldsymbol{\mu}}) \right), \tag{3.22}$$

where n is the number of ensemble members, and $\mathbf{z}_i$ is a vector field of climate model outputs. Furthermore, to evaluate the drawn samples from the distribution, in Chapter 6 we used a classical geostatistical tool of the experimental semi-variogram plots. For additional details of the semi-variogram plots see [102].

In this chapter, we discussed relevant concepts from the geostatistical literature for estimation and simulation. In the next several chapters, we propose a new paradigm for using a high-dimensional manifold for the geostatistical models.

Chapter 4

Non-stationtionary Geostatistical Model

The lines are strong and the colors are diffused . . . as if the artist was in two mind while creating his masterpiece. Similarly, the many choices a mind of a modeler faces . . . is the true process local, regional, or global?

4.1 Introduction

A stochastic process representing complex physical systems is known to exhibit distinct *regional* variability. For example, the sea-level change phenomena near Northern Europe is dominated by the physics of the Glacial-isostatic adjustment. On the other hand, tectonics dominate sea-level changes near Japan [109]. Similarly, topographical characteristics, such as land and ocean, have distinct effects on the regional variability of many climate-related processes.

In Section 1.3, we briefly discussed the problem of estimating the regional sea-level changes near North America. An ad-hoc approach for estimating the process, which addresses the regional variability, includes clustering near-by processes based on domain knowledge. For example, Figure 4.1 depicts an ad-hoc method commonly used in the climate community for estimating sea-level changes around North America [34, 31].

A systematic approach to model the regional variability of a physical process could translate to modeling an appropriate covariance structure for the stochastic process. In this chapter, we propose a framework for modeling a covariance structure that learns the underlying distinct regional variability from the dataset. The core idea of our framework is based on, first, modeling the Gaussian process (GP) using the spatially-varying non-stationary covariance function, and, second, modeling the local property of anisotropy using the Riemannian manifold of symmetric positive definite (SPD) matrices. Hence,

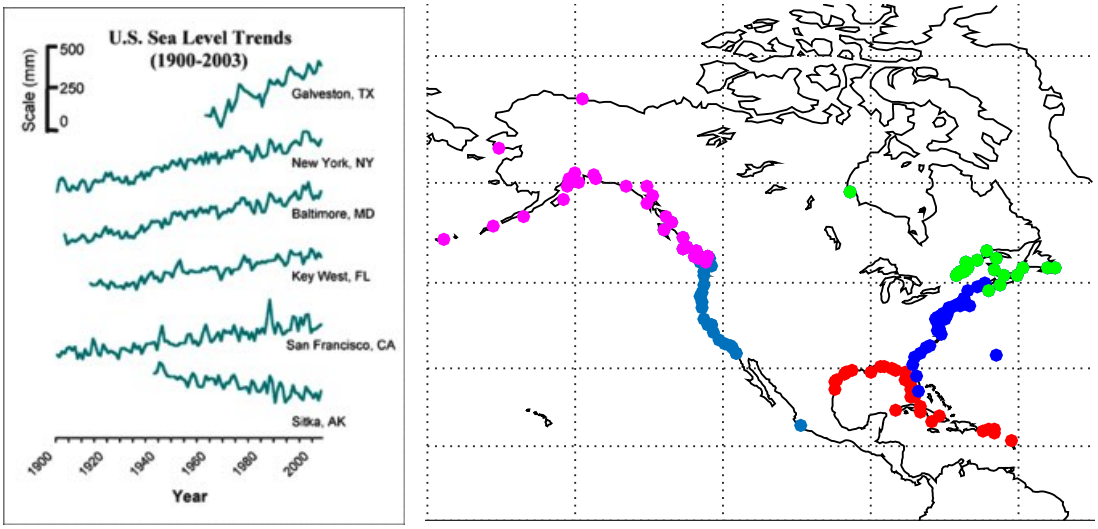


Figure 4.1: Example of a standard climate scientist approach for estimating the spatiotemporal, regional phenomena from climate-related datasets: (a) first, estimate the time series at various tide-gauge sites, and (b) then spatially cluster tide-gauge sites based on geophysical processes to provide the regional phenomena.

we call our framework with the mouth-watering name of the *high-dimensional manifold non-stationary Gaussian process* (HD-NSGP). Relevant concepts for the statistics on SPD spaces is described in Chapter 2, and the background for the spatially-varying non-stationary covariance function is given in Section 3.2. In the next section, we further elaborate on several models in the spatial statistics literature that describe second-order non-stationarity in the process.

4.2 Previous Work

The vast literature on modeling a non-stationary covariance structure, dating as early as the 1990s, can be generally categorized as: spatial deformation models, basis function models, smoothing methods, process convolution models, and stochastic partial differential equation methods.

One of the earliest works on non-stationary models for the environmental application, by [19], used deformation models. The authors used a non-linear mapping to transform the geographic regions to a space which is stationary and isotropic. A Bayesian approach to improve the mapping function of the deformation model was

later introduced by [110, 111]. These Bayesian models require complex and expensive algorithms for parameter estimation; hence, they are usually unfit for high-dimensional spaces. For high-dimensional spaces, up to the dimension of $m = 19$, [112] proposed a meta-modeling formulation using simulation datasets. Another issue with deformation models is that it requires replicates of the dataset to estimate the non-linear mapping function. To address this issue, [113] introduced a deformation model in two-dimensions and from a single realization. The authors partitioned the neighboorhood of the input space into independent processes and estimated parameters using an approximate local-likelihood based approach.

A computationally favorable class of models is the basis function model that was pioneered by Nychka and colleagues [114, 115, 116, 117]. In these models, the central idea is based on approximating the spatial covariance function using a linear (or weighted sum) combination of orthogonal (or non-orthogonal) basis functions. Furthermore, the degree of smoothness in the data was determined by the number of basis function levels used in the model. For large datasets on a regular grid, [115] devised a multiresolution wavelet (non-orthogonal) basis function decomposition. The authors of [115] showed the applicability of their method on the replicates of ozone model outputs around Illinois and Ohio. Later, [117] provided an extension for irregularly spaced input space.

An alternative modeling strategy is to use the Gaussian Markov random field (GMRF) of [118]. GMRF has shown promise for large, discretely indexed datasets. Furthermore, [118] showed an explicit link between GMRF and some classes of non-stationary GPs by solving stochastic partial differential equations (SPDE). The computational benefit was achieved by modeling directly the precision matrix using the GMRF representation. Yet, for applications the overhead cost of problem specific neighborhood structure using triangulation and preprocessing of SPDE has been challenging.

A popular class of models, specifically, process convolution models, naturally construct the local properties by spatially varying the parameters. In these models, and as introduced in [119], the non-stationarity in the process is achieved by using kernel convolution, where the kernel function can be parameterized locally. An alternative

approach by [120, 121, 122] varies the stochastic process over the input space, but, parameterizes a constant kernel function. The resulting convolution of the locally weighted stationary process, in practice, only approximates the chosen covariance function.

In Section 3.2.2, we described the concepts of a particular kind of process convolution model that is most relevant to this thesis. For this model, [7] introduced an analytical solution to process convolution by selecting a Gaussian kernel function and bi-variate Gaussian scale matrix parameter that is spatially-varying. [8] generalized this analytical solution to include the Matérn function and provided a flexible model for the spatially-varying scale matrix parameter using stationary GPs. In [18], an extension to include the spatially-varying smoothness parameter of the Matérn function was formulated. Further extension to multivariate spatial processes when multiple realizations are available was derived in [123]. The spatially-varying non-stationary cross-correlation function, as formulated in [123], is modeled using a kernel smoothed empirical covariance function. For the application of the model in [123], a bi-variate process of temperature and precipitation is estimated from the regional climate model outputs.

A critical bottleneck in the process convolution class of models is estimating the spatially-varying parameters. For large datasets, such as satellite data, [124] designed a Bayesian model for estimating the spatially-varying parameters using low-rank models. [95] shows the limitations of low-rank models for spatial processes and recommends using independent contiguous blocks of input space when modeling. Interestingly, this recommendation aligns with our philosophy of modeling the spatially-varying parameters by establishing a neighborhood structure on a high-dimensional geometric space, and, hence, rendering a flexible regional model.

An alternative strategy by [107, 108, 125] uses covariate information to model the spatially-varying parameters in the non-stationary covariance function. In particular, [107] uses a covariate-based regression for modeling the spatially-varying parameter. Moreover, [107] parameterizes the dependence of the covariates to the spatially-varying scale matrix by constraining the two-dimensional SPD matrix to a three-dimensional conic geometry, and, then, estimating each of the parameters using appropriate priors

in the fully Bayesian fashion. [107] further illustrates the importance of covariate information, such as elevation and its gradient, for modeling the non-stationary process of the two-dimensional precipitation dataset around Colorado.

In the next section, we propose an unconstrained, non-parametric, data-driven model for the spatially-varying scale matrix parameter of the non-stationary covariance function. We further discuss the potential of including the covariate information or a data-fusion model in Chapter 5. The main idea for our model is to use the property of SPD space and to establish the high-dimensional geometric structure using the Riemannian manifold framework. Hence, our approach allows for additional flexibility in the degree of smoothness of the scale matrix parameter and guarantees a valid non-stationary covariance function. The relevant background for estimating SPD matrices on a Riemannian manifold is described in Chapter 2.

In this section we discussed various non-parametric methods for modeling non-stationarity in the spatial processes for a particular domain of applications. None, including our model, is a one-size-fit-all application. Furthermore, most of the non-stationary methods require complex model-fitting procedures, and, for an implementation in a general setting, a readily available software has not yet been provided. Hence, most methods, including ours, will usually resort to comparing with stationary models for which a general software is readily available.

4.3 High-Dimensional Manifold Non-stationary Gaussian Process

An overview of our framework for modeling the spatially-varying geometric anisotropy, or the scale matrix, for the non-stationary covariance function is illustrated in Figure 4.2. Here, the fit of the ellipses around the tide-gauge locations are improved by using intrinsic statistics on the high-dimensional manifold of SPD matrices.[1]

[1]In this chapter, we represent the points in the input space with the notation $\boldsymbol{x_i} \triangleq \boldsymbol{x(s_i)}$ so that, later, we can generalize the model for the implementation of spatiotemporal datasets or the covariate information.

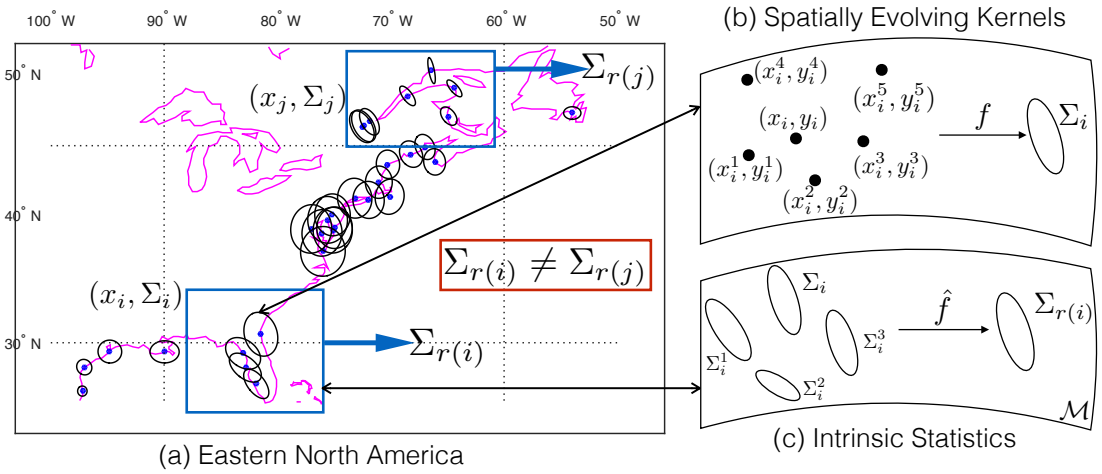


Figure 4.2: The proposed framework for the high-dimensional-manifold non-stationary covariance function. Here, the geometric anisotropy (i.e., ellipses) around (a) eastern North America are modeled using the (b) spatially evolving kernels and (c) further improved by estimating them on a statistical manifold of symmetric positive definite matrices.

4.3.1 Proposed Model

The model for the stochastic process, $\{Z(\boldsymbol{x}) : \boldsymbol{x_i} \subset \mathbb{R}^m\}$, using the Gaussian process is given in Equation 3.2, and for the spatially-varying non-stationary Matérn correlation function, $\rho^{NS}(\cdot, \cdot)$, see Equation 3.12. An important ingredient in the correlation function, besides the smoothness parameter, is what we call the dissimilarity function, $q^{NS}(\boldsymbol{x_i}, \boldsymbol{x_j})$. In Equation 3.13, a model of the dissimilarity function by [8] is outlined.[2] Here, the spatially-varying geometric anisotropy, Σ_{ij}, is an average of the scale matrices, $\{\Sigma_i, \Sigma_j\}$. The scale matrices are centered at each location in the input space, $\boldsymbol{x_i}$, and are modeled using an eigendecomposition, as described in Equation 3.17. We propose an alternative strategy to model the dissimilarity function, and, consequently, the spatially-varying non-stationary covariance function, as follows:

[2]In [8], the quadratic form of q_{ij}^{NS} is also called the scale mixture of kernels, and the scale matrix, Σ_i, is called the kernel matrix.

$$\rho^{HDNS}(\boldsymbol{x_i}, \boldsymbol{x_j}) = 2^{\frac{\mathrm{m}}{2}} |\Sigma_{r(i)}|^{\frac{1}{4}} |\Sigma_{r(j)}|^{\frac{1}{4}} \left|\frac{\Sigma_{r(i)}+\Sigma_{r(j)}}{2}\right|^{-\frac{1}{2}} \rho^{S}(\nu, q_{ij}^{HDNS}), \tag{4.1a}$$

$$q_{ij}^{HDNS} = (\boldsymbol{x_i} - \boldsymbol{x_j})^T \left(\frac{\Sigma_{r(i)}+\Sigma_{r(j)}}{2}\right)^{-1} (\boldsymbol{x_i} - \boldsymbol{x_j}), \tag{4.1b}$$

$$\Sigma_{r(i)} \sim \mathcal{N}(\bar{\Sigma}, \Lambda). \tag{4.1c}$$

Given a region, $r(i) \triangleq r(\boldsymbol{x_i})$, associated with the location, $\boldsymbol{x_i}$, the scale matrix parameter, $\Sigma_{r(i)}$, is modeled using the intrinsic statistics on a manifold of SPD matrices, $S^+(m, \mathbb{R})$. In Equation 2.5, we describe the probability density function (PDF) of the generalized normal distribution for the SPD matrices with the a mean parameter, $\bar{\Sigma}$, and a variance-covariance parameter, Λ, as defined on the Riemannian manifold. To circumvent the computationally intensive normalizing factor of the PDF, Equation 2.6 provides a simplified version. The PDF in Equation 2.7 describes further simplification by replacing the variance-covariance matrix, Λ, with a scalar variance, σ^2, version of the generalized normal distribution.

The primary difference in our covariance function construction, as described above, when compared to Equation 3.12, as given in [8], is the model for the scale matrix, Σ_i. While [8] prescribes stationary GPs for each of the parameters in the eigendecomposition of the scale matrix, we directly sample, or estimate, the scale matrix from the distribution defined on the manifold of SPD matrices. However, we still need to show that our construction is valid. In the theorems below, we show the validity of our construction.

Theorem 1. *The spatially-varying non-stationary covariance function, as constructed from the Equation 4.1, is a valid non-stationary covariance function.*

Proof. For the covariance function to be valid the only requirement is that it is positive definite. Theorem 1 in [8], states that the construction of the non-stationary correlation function of Equation 4.1a is positive definite as long as the scale matrix associated to each location is a SPD matrix, i.e., $\Sigma_i \in S^+(m, \mathbb{R})$.

Our formulation of the scale matrix, $\Sigma_{r(i)}$, at each location or its associated region, as given in Equation 4.1b, is constructed on a smooth, i.e., C^{∞}, Riemannian manifold of SPD matrices, $\Sigma_{r(i)} \in \mathrm{GL}(m, \mathbb{R})$. Furthermore, we chose an affine-invariant metric, and its associated affine connection, to facilitate the algebraic operations in $\mathrm{gl}(m, \mathbb{R})$ for the estimation, or sampling, of the real symmetric matrices, $\mathrm{S}(m, \mathbb{R})$, on the tangent space of the manifold, $\mathcal{T}_{\bar{\Sigma}}\mathcal{M}$. In particular, the exponential map, $\mathrm{Exp}_{\bar{\Sigma}}(\Sigma_{r(i)}) : \mathrm{S}^{+} \rightarrow \mathrm{S}$, and the logarithmic map, $\mathrm{Log}_{\bar{\Sigma}}(\Sigma_{r(i)}) : \mathrm{S} \rightarrow \mathrm{S}^{+}$, provides us with essential tools to perform the intrinsic statistics for scale matrices, $\{\Sigma_{r(i)}\}$, and, further, guarantee that $\Sigma_{r(i)} \in \mathrm{S}^{+}(m, \mathbb{R})$. For details, see Chapter 2 and the references of the proofs therein mentioned.

By construction, on a smooth manifold with the affine-invariant metric we are guaranteed an outcome of the scale matrix estimation procedure to be a SPD matrix, i.e., $\Sigma_{r(i)} \in \mathrm{S}^{+}(m, \mathbb{R})$. Hence, our prescription of the non-stationary covariance function, as given in Equation 4.1, is valid.

□

Theorem 2. *A non-stationary Gaussian process, $Z(\cdot)$, is Mth-order sample path and mean square differentiable for the non-stationary correlation function, $\rho^{HDNS}(\cdot, \cdot)$, as given in Equation 4.1, if the associated stationary correlation function, $\rho^{S}(\cdot)$, is Mth-order sample path and mean square differentiable.*

Proof. The proof is an application of the infinitely differentiable, complete Riemannian metric space of the scale matrix parameter, $\Sigma_{r(i)}$, to the sample path and mean square differentiability theorems of the non-stationary Gaussian process in [8] (i.e., Theorem 11 and Theorem 12).

We assume without loss of generality that the signal variance, $\sigma_Z^2(\boldsymbol{x})$, and the noise variance, $\sigma_Y^2(\boldsymbol{x})$, as given in the covariance function formulation of Equation 3.7, is either constant or infinitely differentiable. Hence, the smoothness properties of the process, in turn, primarily depends only on the mean square and sample path properties of the correlation function.

Moreover, the smoothness properties of the process in the theorems of [8], have

an additional dependence on the smoothness properties of the independent stationary Gaussian processes involved in modeling the scale matrix parameter. In particular, the scale matrices are non-singular and its elements suffice at least M+1th-order sample path differentiability and Mth-order mean square differentiability. In Lemma 8 (of [8]), these conditions on their model of the scale matrix parameter are shown to exist.

In our model of the scale matrix parameter, the manifold chosen is C^{∞} differentiable. Hence, we are guaranteed at least an M+1th-order partial derivative of the random fields, as given in Equation 2.5, and, that is associated to the scale matrices, i.e., $\Sigma_{r(i)}$, $\Sigma_{r(j)}$, and $(\Sigma_{r(i)} + \Sigma_{r(j)})$. Furthermore, the chosen metric and the distance function on the manifold, as given in Section 2.2.3 and Equation 2.3, is complete. Hence, we are guaranteed that the scale matrices are bounded and non-singular. Therefore, by construction the smoothness properties of sample path and mean square differentiability for the non-stationary Gaussian process only depends on the associated, chosen stationary correlation function. Note, the smoothness properties of the stationary correlation function are described in Section 3.2.2.

□

Hence, by introducing intrinsic statistics directly on the SPD space of the scale matrix parameter, rather than a separate stationary process for each element of the matrix, one can potentially achieve greater flexibility in its estimation while maintaining the smoothness properties of the primary spatial process, $\{Z(\cdot)\}$, used for modeling the data. In the next section, we describe the techniques for the parameter estimation of our model.

4.3.2 Parameter Estimation

For complex, high-dimensional models, including our proposed model from the previous section, parameter estimation using likelihood-based schemes is usually analytically and numerically intractable. In Section 3.2.1, we pointed out the widely used parameter estimation scheme for a GP model, which is the Markov Chain Monte Carlo (MCMC)

method. For Bayesian inference, MCMC generates samples of the posteriors, or predictive distributions, when the chain reaches the equilibrium distribution. In other words, the goal is to sample the posteriors of the parameters, $p(\theta|\{\mathbf{Y}, \mathbf{x}\})$, from the training set, $\{\mathbf{Y}, \mathbf{x}\}$, such that the posterior predictive mean is $\widehat{E[\mathbf{Y}^*|\mathbf{Y}]} = L^{-1} \sum_{l=1}^{L} \mathbf{Y}_l^*$. Here, the posterior samples, $\mathbf{Y}_l^*$, at the test location, $\mathbf{x}^*$, are generated from the conditional distribution of $[\mathbf{Y}_l^*|\theta_l, \mathbf{x}^*, \mathbf{Y}, \mathbf{x}]$, as described in Equations 3.4 and 3.5.

In the MCMC literature, there are vast variants of the sampling schemes developed for certain classes of models or applications [126, 127]. For example, the NSGP model in [8] prescribes a posterior mean centering scheme (PMC) to propose the posteriors and the parameters of the model jointly. Additionally, PMC was designed to deal with the various mixing issues of the embedded GP processes in the NSGP model. Another popular variant of MCMC in the high-dimensional covariate space is a hybrid scheme, such as the Hamiltonian Monte Carlo scheme (HMC) [128] or the Metropolis Adjusted Langevin Algorithm (MALA) [129]. These hybrid schemes are useful because they provide the appropriate direction of the proposals toward the target densities by using the information of the gradient of the process. However, it is not clear if the hybrid models would improve the hyperparameter mixing of complex non-stationary models.

For our proposed model, the gradient of the process in closed form is currently not available. Furthermore, the primary goal of the variants of the basic MCMC scheme is to improve the mixing of the chain, and, moreover, to aid its convergence towards the target distribution. Hence, for our pilot implementation, we resort to a flexible MCMC scheme by [130], which we have found to provide satisfactory results in our simulation study in Section 4.4. In addition, this MCMC scheme is also applicable to non-Gaussian data. Intuitively, the implemented MCMC scheme guides the joint proposals by using an elliptical slice sampler in conjunction with a surrogate GP. The advantage of using a surrogate GP step is that it randomly reparameterizes the proposals around the current state. Also, the elliptical slice sampler adaptively adjusts the hyperparameters of the covariance function.

For the proposed covariance function, i.e., Equation 4.1, the parameters to estimate are the signal variance, noise variance, smoothness, and scale matrices, i.e.,

$\theta_Y = \{\sigma_Y, \sigma_Z, \nu, \Sigma_i\}$. In addition, the hyperparameters for the scale matrix model, $\{\Sigma_i\}$, are $\theta_\Sigma = \{|r|, \{\bar{\Sigma}\}_{(1,\ldots,|r|)}, \{\Lambda\}_{(1,\ldots,|r|)}\}$. Here, the number of clusters, or regions, are $|r|$, and, the mean and variance-covariance matrices for the regional distribution of the scale matrix on a manifold are $\bar{\Sigma}_{r(x_i)}$ and $\Lambda_{r(x_i)}$, respectively. The total number of parameters to estimate at the first level in the model hierarchy is $|\theta_{\text{level}=1}| \triangleq |\theta_Y| = 3 + |r| \cdot (m + \frac{m(m-1)}{2})$, and at the second level is $|\theta_{\text{level}=2}| \triangleq |\theta_\Sigma| = 1 + |r| \cdot (m + \frac{m(m-1)}{2}) + |r| \cdot (m^2 + \frac{m^2(m^2-1)}{2})$. Hence, the number of parameters to be estimated, $|\theta_{\text{est}}|$, discounting the parameters of the priors for $\theta_{\text{level}=1}$ and $\theta_{\text{level}=2}$, can be written as follows:

$$|\theta_{\text{est}}| = \Big(3\Big) + \Big(1 + |r| \cdot (m + \tfrac{m(m-1)}{2}) + |r| \cdot (m^2 + \tfrac{m^2(m^2-1)}{2})\Big). \tag{4.2}$$

Note, the dimension for the parameter space, discounting the prior, is $|\theta_{\text{total}}| = |\theta_{\text{est}}|$.

The computational complexity for generating each sample in the MCMC involves $\mathcal{O}(n^3)$ for the first level and $\mathcal{O}(|r| \cdot (m^2)^3)$ for the second level. The hyperparameters for the scale matrix can be further reduced by assuming a diagonal only, or scalar, variance and a homogeneous, or stationary, scale matrix process within the regions, i.e., $\theta_\Sigma = \{|r|, \bar{\Sigma}, \boldsymbol{\sigma}_\Sigma\}$. This reduces the number of parameters in the model to $|\theta_\Sigma| = 1 + (m + \frac{m(m-1)}{2}) + m^2$ and the computations complexity for the second layer to $\mathcal{O}((m^2)^3)$. Hence, the number of parameters to be estimated, $|\theta_{\text{est}}|$, can be written as follows:

$$|\theta_{\text{est}}| = \Big(3\Big) + \Big(1 + \cdot(m + \tfrac{m(m-1)}{2}) + m^2\Big). \tag{4.3}$$

Here, the dimension for the parameter space, discounting the prior, is as follows: $|\theta_{\text{total}}| = (|\theta_{\text{est}}| - 3) \cdot |r| + 3$.

In Proposal Scheme 1, we outline the suggested MCMC scheme for our proposed GP model (HD-NSGP). Steps 1 and 2 in the MCMC scheme refer to the surrogate GP step; Steps 3, 11, and 13 refer to the elliptical slice sampler steps; and Step 4 refers to the parameter proposal. The parameter proposals corresponding to our covariance model are further outlined in Proposal Schemes 2 and 3. Note, without loss of generality and to simplify the prior specification, we assume the input space to lie in an m-dimensional

unit hypercube, i.e., $\boldsymbol{x_i} \in [0,1]^m$.

Proposal Scheme 1: Outline of the suggested MCMC scheme for the HD-NSGP model.

Input: GP model (from Equation 3.2), covariance function (from Equation 4.1), vector of response or latent variables, $\mathbf{Y}$, initialized parameter values of the model, $\theta = \{\theta_Y, \theta_\Sigma\}$, vector of elliptical slice sampler scale, $\boldsymbol{\sigma}_{\text{ess}}$.

Output: Posterior samples of θ', $\mathbf{Y}'$.

1: Sample surrogate data $\mathbf{h} \sim \mathcal{N}(\mathbf{Y}, K_{\text{surr}}(\theta))$.

2: Compute implied variates $\boldsymbol{\omega} = L^{-1}(\theta)(\mathbf{Y} - \boldsymbol{\mu}_{\theta,\mathbf{h}})$. The cholesky decomposition over here refers to the auxiliary GP model's covariance matrix, i.e., $L(\theta)L^T(\theta) = (K^{-1}_{\text{hd-nsgp}}(\theta) + K^{-1}_{\text{surr}}(\theta))^{-1}$.

3: Re-center the parameter bounds: $\mathbf{b} \sim \mathcal{U}(\mathbf{0}, \boldsymbol{\sigma}_{\text{ess}}), \theta_{\min} = \theta - \mathbf{b}, \theta_{\max} + \boldsymbol{\sigma}_{\text{ess}}$.

4: Propose new sample of parameters from the proposal distribution, i.e., $\theta' \sim q(\theta'; \theta, \theta_{\min}, \theta_{\max})$. There are two parts to this proposal. First, is the proposal of θ_Y, which is explained in Proposal Scheme 2. Second, is the proposal of θ_Σ, which is described in Proposal Scheme 3.

5: The Metropolis-Hasting step, i.e., draw $a \sim \mathcal{U}(0,1)$, where the Metropolis-Hasting's ratio is as follows: $\min\left(1, \frac{P(\theta', \mathbf{Y}'|\text{data}) \cdot P(\mathbf{h}|\mathbf{Y}', K_{\text{surr}}(\theta')) \cdot q(\theta;\theta') \cdot |L^{-1}(\theta)|}{P(\theta, \mathbf{Y}|\text{data}) \cdot P(\mathbf{h}|\mathbf{Y}, K_{\text{surr}}(\theta)) \cdot q(\theta';\theta) \cdot |L^{-1}(\theta')|}\right)$.

6: Determine the threshold for the slice sampler: $\tau_{\text{ess}} = a\mathcal{L}(\mathbf{Y})\mathcal{N}(\mathbf{h}; \mathbf{0}, K_{\text{hd-nsgp}}(\theta) + K_{\text{surr}}(\theta))p(\theta)$.

7: Compute the posterior samples $\mathbf{Y}' = L(\theta')\boldsymbol{\omega} + \boldsymbol{\mu}_{\theta',\mathbf{h}}$.

8: **if** $(\mathcal{L}(\mathbf{Y}')\mathcal{N}(\mathbf{h}; \mathbf{0}, K_{\text{hd-nsgp}}(\theta') + K_{\text{surr}}(\theta'))p(\theta') > \tau_{\text{ess}})$ **then**

9: **return** $\{\mathbf{Y}', \theta'\}$.

10: **else if** $(\theta' < \theta)$ **then**

11: shrink the proposal bracket by $\theta_{\min} = \theta'$.

12: **else**

13: shrink the proposal bracket by $\theta_{\max} = \theta'$.

14: **end if**

15: Loop back to step 4.

The surrogate GP augments the response variable, $\mathbf{Y}$, with an auxiliary variable, $\mathbf{h}$, such that $[\mathbf{h}|\mathbf{Y},\theta] \sim \mathcal{N}(\mathbf{Y}, K_{\text{surr}}(\theta))$. The marginal distribution is then given as $[\mathbf{h}|\theta] \sim \mathcal{N}(\mathbf{0}, (K_{\text{hd-nsgp}} + K_{\text{surr}}))$, and the posteriors are $[\mathbf{Y}|\mathbf{h},\theta] \sim \mathcal{N}(\mu_{\mathbf{h},\theta}, (K^{-1}_{\text{hd-nsgp}} + K^{-1}_{\text{surr}})^{-1})$. By using basic linear algebra, the auxiliary mean values are $\mu_{\mathbf{h},\theta} = K_{\text{hd-nsgp}}(K_{\text{hd-nsgp}} + K_{\text{surr}})^{-1}\mathbf{h}$. For convention, we denote the covariance matrix of our proposed model as $K_{\text{hd-nsgp}}$, the covariance matrix of the surrogate GP as K_{surr}, and the conditional likelihood as $\mathcal{L}(Y) = P(\text{training data}|Y)$. A straightforward surrogate covariance model of $K_{\text{surr}}(\theta) = \alpha\mathbf{I}$ acts as a free parameter that limits the sampling region for the updates of Y. The NSGP model in [8] uses a similar reparametrization scheme by centering the sampling updates around the posterior means. An alternative fixed reparametrization scheme is also described in [131]. Note, when $\alpha \to \infty$, the MCMC scheme tends to uninformative priors in the sampler, and, in the limiting cases, it acts similar to that of whitening the priors.

For the analytical GP model, one can model the surrogate covariance using a Laplacian approximation. Unfortunately, our GP model does not have a closed-form solution; hence, we chose an appropriate α based on the preliminary runs. Another free parameter in Scheme 1 is the slice sampling parameter, σ_{ess}, which adjusts the width of the prior distribution. In our experiments, we appropriately set the width, σ_{ess}, to the largest feasible values of the priors. Furthermore, the priors were assumed to be independent, i.e., $p(\theta) = p(\theta_Y)p(\theta_\Sigma)$, and all of the hyperparameters for the prior distribution were chosen so that they were vague but proper.

In Proposal Scheme 2, we outline the sampling procedure for the geostatistical parameters of the signal variance, noise variance, smoothness parameter, and scale matrix. The smoothness parameter of the Matérn covariance function is known to exhibit difficulties in the estimation. For the technical details on estimating the smoothness parameter see [132]. For the climate-related dataset, it is well known that $\nu = 1$ provides a good fit of the underlying smoothness of the process. Hence we used the profile likelihood scheme, or integrated the smoothness parameter out of the model using a discrete set of parameters, $\nu = \{0.5, 1, \ldots, 5, 5.5\}$. For further details on the profile likelihood scheme see [1]. Other parameters in the Matérn function, i.e., signal and noise variances,

were proposed from their conditional distributions and using appropriate conjugate priors. Furthermore, the hyperparameters of the priors were re-centered from Step 3 of Scheme 1. Alternatively, many authors have recommended a re-parameterization of the variances for improving the estimation. For example, [133] proposes to reparameterize the geostatistical parameters of the Matérn function to $\{\log(\sigma_Z^2/\Sigma^{2\nu}), \log(\Sigma^{2\nu})\}$.

Proposal Scheme 2: For the geostatistical parameters, $\theta_Y = \{\sigma_Z, \sigma_Y, \nu, \Sigma\}$.

1: For the priors of ν, a small sample of discrete set were used, i.e., $\nu = \{0.5, 1, \ldots, 5, 5.5\}$.
2: Propose noise variance, σ_Y^2, from their conditionals $[\sigma_Y|\mathbf{Y}, \sigma_Z, \nu, \Sigma]$ and conjugate priors $p(\sigma_Y) \sim \mathcal{IG}(a_Y, b_Y)$.
3: Propose signal variance, σ_Z^2, from their conditionals $[\sigma_Z|\mathbf{Y}, \sigma_Y, \nu, \Sigma]$ and conjugate priors $p(\sigma_Z) \sim \mathcal{IG}(a_Z, b_Z)$.
4: Compute the scale matrix, $\Sigma_{r(i)}$, from the Proposal Scheme 3 or 8.

Proposal Scheme 3 outlines the sampling procedure for estimating the scale matrix parameter. In Step 1, the nearest neighborhood parameter, $|r|$, of the spatially-varying scale matrix, $\Sigma_{r(i)}$, is sampled from a discrete set of values. For the climate-related dataset, we chose an appropriate range of values for $|r|$ that are informed by the geophysics of the climate variable. Alternatively, one can parametrize the radius of the spatial clusters from the measure of dispersion in the scale matrix samples, i.e., $\text{tr}(\Lambda(\Sigma_{i=1}, \ldots, \Sigma_{|r|})) \leq \tau_{|r|}^2$, where the threshold is proposed from the priors, i.e $\tau_{|r|} \sim \mathcal{N}(0, 1)$. In this scheme, the spatially-varying scale matrix is set to the mean of the spatial cluster, i.e., $\Sigma_{r(i)} = \bar{\Sigma}(\Sigma_{i=1}, \ldots, \Sigma_{|r|})$. Here the mean and variance of the scale matrix are computed on the SPD manifold using Equations 2.9 and 2.12, respectively. See [134], for the recommendation of using the trace of the diffusion tensor, Λ, as the direct measure of dispersion in the local diffusion process. The intuition for the aforementioned proposal is that small differences in the scale matrix within a region could possibly result from an artifact of the estimation procedure, an over-smoothed parameterization, or small datasets. In such cases, a conservative modeling approach

Proposal Scheme 3: For the spatially-varying geometric anisotropy or scale matrix parameters, $\theta_\Sigma = \{|r|, \bar{\Sigma}, \Lambda\}$.

1: Sample $|r| \sim \mathcal{U}\{a_{|r|}, b_{|r|}\}$, where $\min(a_{|r|}) = 1$ and $\max(b_{|r|}) = n$. In the case of climate-related datasets, the hyperparameters are informed by the geophysics of the climate variable. In the case of simulated datasets, we use uninformative priors, such as the uniform discrete distribution.

2: To jointly sample $\Sigma_{r(i)} \sim \mathcal{N}(\bar{\Sigma}, \Lambda)$, we can use the spectral decomposition property of the SPD matrix, such that $\Sigma_{r(i)} = \hat{\Sigma}(\lambda, U) \cdot \bar{\Sigma}^{1/2} \sim \mathcal{N}(\bar{\Sigma}, \sigma^2_\Sigma \mathbf{I})$ and $\hat{\Sigma}(\lambda, U) \sim \mathcal{N}(\mathbf{I}, \sigma^2_\Sigma \mathbf{I})$.

3: To sample U, first sample $\{A_i\}_{(i=1,\ldots,m\times m)}$ independently from $\mathcal{N}(0,1)$ to form a matrix of $A_{m\times m}$. Then, use QR-decomposition to find U, i.e., $[U, \sim] = \text{qr}(A)$. The theorems for such a construction can be found in [22](pp. 70).

4: To sample $\lambda = \{\lambda_1, \ldots, \lambda_m\}$ jointly from the multivariate PDF of Equation 2.7, we follow the construction of [92], as given below:
$p(\lambda) = (m!2^m)^{-1}\, \omega_m\, 8^{(m(m-1)/4)}\, k^{-1}_{scG}(\sigma_\Sigma)\, e^{(\lambda_1^2+\ldots+\lambda_m^2)/2\sigma^2_\Sigma} \prod_{i<j} \sinh(|\lambda_i - \lambda_j|/2)$,
where $\omega_m = \frac{2^m \pi^{(m^2/2)}}{\Gamma(m/2)}$ and k_{scG} is given in Equation 2.12. When $m = 2$, the above PDF simplifies to an analytical solution for $\{\lambda_1, \lambda_2\}$, and when $m \geq 3$, a look-up table can be generated.

5: Finally, to sample $\{\Sigma_{r(i)}\}$, compute the following: $\Sigma_{r(i)} = \hat{\Sigma} \cdot \bar{\Sigma}^{1/2}$.

of assuming stationarity within a small region could potentially produce an improved prediction.

Another alternative to the HD-NSGP model is first to train the model using the NSGP model, and then to represent the spatially-varying scale matrix using the cluster mean. This modeling strategy is briefly descried in Proposal Scheme 8, Appendix B.2.1. Here, the cluster size is estimated using the aforementioned proposal scheme, i.e., either informed by the geophysics or estimated by constraining the scale matrix dispersion within the spatial regions. Such a model allows the flexibility of regional stationarity to the spatially-varying NSGP, i.e., $\Sigma_i \to \Sigma_{r(i)}$. For example, when $|r| = 0$ the NSGP

models reduces to the SGP model; when $|r| = n$, the model is NSGP; and when $0 < |r| < n$, the NSGP model follows the HD-NSGP model. In such a modeling strategy, the training stage uses the additional parameterization of the GP processes for estimating the scale matrix, but the predictive, or posterior, distribution still follow the proposed HD-NSGP model.

To avoid using additional GP processes in the training phase, we sample the scale matrix directly from the SPD manifold, i.e., $\Sigma_{r(i)} \sim \mathcal{M}(\bar{\Sigma}, \Lambda)$. The samples for $\Sigma_{r(i)}$ can either follow the PDF of Equation 2.5 or the simplified normalization factor version as given in Equation 2.6. For computational simplicity, the PDF with a scalar version of the variance parameter, and for $m = 2$, is given in Equation 2.7. The proposal of scale matrix parameters, $\{\bar{\Sigma}, \sigma_{\Sigma}^2 \mathbf{I}\}$, are outlined in Steps 2 to 5 of Proposal Scheme 3. The joint proposal of the mean, $\bar{\Sigma}$, and variance, σ_{Σ}, follows from the construction described in [22] and [92]. A straightforward extension to a variance-covariance matrix, i.e., $\sigma_{\Sigma} \to \Lambda$ with $m \geq 2$ can similarly be constructed. However, the computational complexity increases as the normalization factors, k_{G} and k_{SG}, are more involved in the PDF of Equations 2.5 and 2.6, respectively. For an extension to $m \geq 2$, a lookup table for the normalizing factor, k_{scG}, can be generated using the Monte Carlo method.

We initialize the sampler using the parameters of the SGP model fit. The scale matrix at the test points is sampled from its association to the nearest neighbor training data points. Note, since the conditionals of the scale matrix parameter during the joint proposal is unknown in Proposal Scheme 3, they are sampled from the prior correlation instead of the more favorable choice of the posterior correlation. Furthermore, the priors for the simulation dataset are chosen uninformatively, but, for the climate-related datasets, informative priors improve the estimation. Hence, for the sea-level datasets, we follow the priors and their hyperparameter values from [34]. A useful future exercise would be to develop a maximum likelihood-based estimation procedure for HD-NSGP, similar to [113], for improving the estimation procedure.

4.3.3 Evaluation Procedure

The stationary and non-stationary GP run on the test set, $\{\boldsymbol{x_i^*}, y_i^*\}_{i=1,\ldots,n}$, was evaluated using the following three criteria: the mean squared error (MSE), negative log predictive density (NLPD), and continuous rank probability score (CRPS).

The mean square error evaluates the predictive mean estimates, $\{\hat{y_i}\}$, of the model at the test locations using the following formulation:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i^* - \hat{y}_i)^2. \tag{4.4}$$

For the GP model, the posterior predictive mean estimate was averaged over the MCMC sample, $\{\theta_j\}_{j=1\ldots J}$, draws, $\boldsymbol{Y^*}|\boldsymbol{Y}, \theta_j$, from the posterior $p(\theta|\boldsymbol{Y})$.[3] A smaller MSE value suggests a better point estimates of the dataset.

To compare the entire predictive distribution under consideration, we evaluated the negative log predictive density, $\text{NLPD} = -\log(\hat{p}(\boldsymbol{Y^*}|\boldsymbol{Y}))$. NLPD criteria is discussed in [135] and was used by [8, 107] to evaluate non-stationary models. For the GP model, NLPD was averaged over the MCMC draws from the posteriors as follows:

$$\text{NLPD} = -\log\left(\frac{1}{J}\sum_{j=1}^{J} p(\boldsymbol{Y^*}|\boldsymbol{Y}, \theta_j)\right). \tag{4.5}$$

A smaller NLPD value suggests a better fit of the predictive model.

Finally, as pointed out in [107], we used the continuous rank probability score, instead of the information-theoretic criteria, to compare the stationary and non-stationary models. Information-theoretic criteria, such as KL-divergence, was previously used by [8] to evaluate the non-stationary model. As discussed in [136], stationary models and non-stationary models have different ranks of kernel process representation, and, hence, CRPS is a better metric for evaluation. CRPS is introduced in [137] and defined as: $\text{CRPS}(F_i, y_i^*) = -\int_{-\infty}^{\infty}(F_i(x) - \text{I}(x \geq y_i^*))^2 dx$. Here, $F_i(\cdot)$ is the cumulative distribution function for y_i^*, and $\text{I}(\cdot)$ is the indicator function. For the GP model, $F_i(\cdot)$ is the conditionally Gaussian predictive cumulative distribution function, and CRPS is

[3] See Equation 3.5 for the predictive mean estimate of the GP model.

averaged over the MCMC draws from the posteriors as follows:

$$\text{CRPS} = n^{-1} \sum_{i=1}^{n} \Big(- J^{-1} \sum_{j=1}^{J} \int_{-\infty}^{\infty} \big(F_i(x;\theta_j) - \text{I}(x \geq y_i^*)\big)^2 dx \Big). \tag{4.6}$$

A larger CRPS value suggests a better fit of the predictive model.

4.4 Simulation Study

To gain insight into the applicability of our proposed covariance function, we constructed a simulation function from a regionally parameterized Gaussian process. Our construction of the simulation function was influenced by many of the climate-related processes that are known to exhibit regional characteristics. The proposed simulation function also allowed for various degrees of smoothness within and across the regions. Such smoothness properties in the physical process are commonly found in many climate-related processes. Examples of some of the climate-related processes with regional characteristics are discussed in Section 1.3 and illustrated in Figure 1.1.

In [8], the two-dimensional simulation study for the non-stationary model used the Hwang function [138], which is based on the complex interaction of surfaces. A realization from the Hwang function is shown in Figure 4.3a. Even though this simulation function is non-stationary, it is fairly smooth and homogeneous. Furthermore, it lacks the complex regional geophysics that is commonly encountered in climate-related processes. Hence, in the next section, we propose a simulation function that emulates well the physical process. A realization from our proposed simulation function is shown in Figure 4.3b.

4.4.1 Experimental Setup

The univariate simulation function, $y(\cdot)$, from a two-dimensional input space, $m = 2$, was constructed as follows:

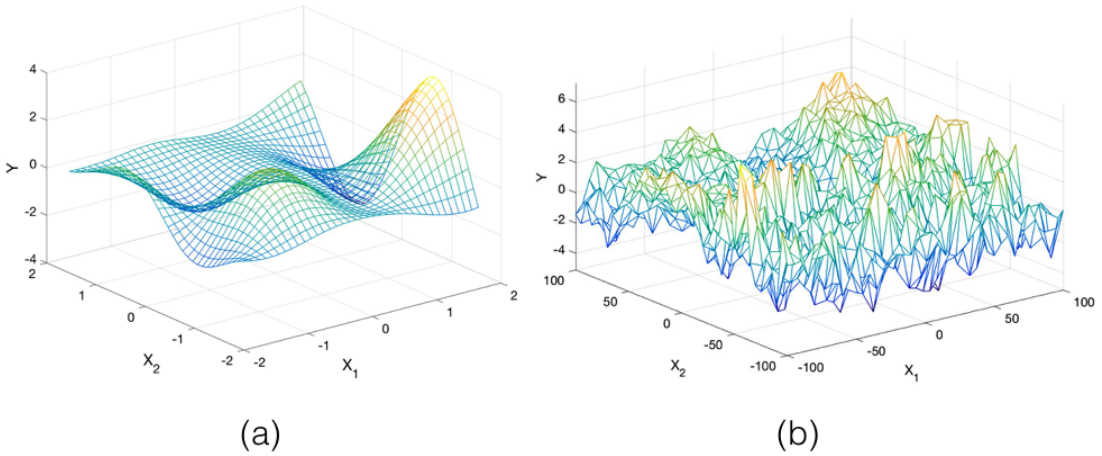


Figure 4.3: Simulated surfaces from the (a) Hwang function and (b) our proposed simulation function.

$$\boldsymbol{y} = f_1(\boldsymbol{x})w(\boldsymbol{x}) + f_2(\boldsymbol{x})(1 - w(\boldsymbol{x})), \tag{4.7a}$$

$$f_i(\boldsymbol{x}) \sim \text{GP}(0,\, \sigma_{f_i}^2 H^S(\boldsymbol{x}; \nu_{f_i}, \Sigma_{f_i}) + \sigma_{y_{f_i}}^2 \mathbf{I}), \tag{4.7b}$$

$$w(\boldsymbol{x}) = A_w + \frac{K_w - A_w}{\left(C_w + e^{-B_w(\boldsymbol{x} - M_w)}\right)^{1/\nu_w}}. \tag{4.7c}$$

Here, the two-dimensional input space was modeled using two separate stationary Gaussian processes, $\{f_i\}_{(i=1,2)}$, with the parameters of signal variance, noise variance, Matérn's smoothness parameter, and scale matrix, $\{\sigma^2, \sigma_y^2, \nu, \Sigma\}_{f_i}$. The stationary correlation function was Matérn, as given in Equation 3.9, and was used to compute the correlation matrix, H^S. Figure 4.3b shows a sample from the realization of our proposed simulation function. Similarly, Figure 4.4a shows a sample when the noise variance was excluded. Based on visual inspection of Figure 4.4a, one can see that the input space regions of $\boldsymbol{X_2} > 0$ and $\boldsymbol{X_2} < 0$ is primarily a Gaussian process realization from the function of f_1 and f_2, respectively.

To control the effects of the underlying stationary GPs, $\{f_i\}$, within the regions and their transitions across the boundaries at $\boldsymbol{X_2} = \mathbf{0}$, we weighted the GPs using a generalized logistic function in two dimensions, $w(\boldsymbol{x})$, as given in Equation 4.7c. The

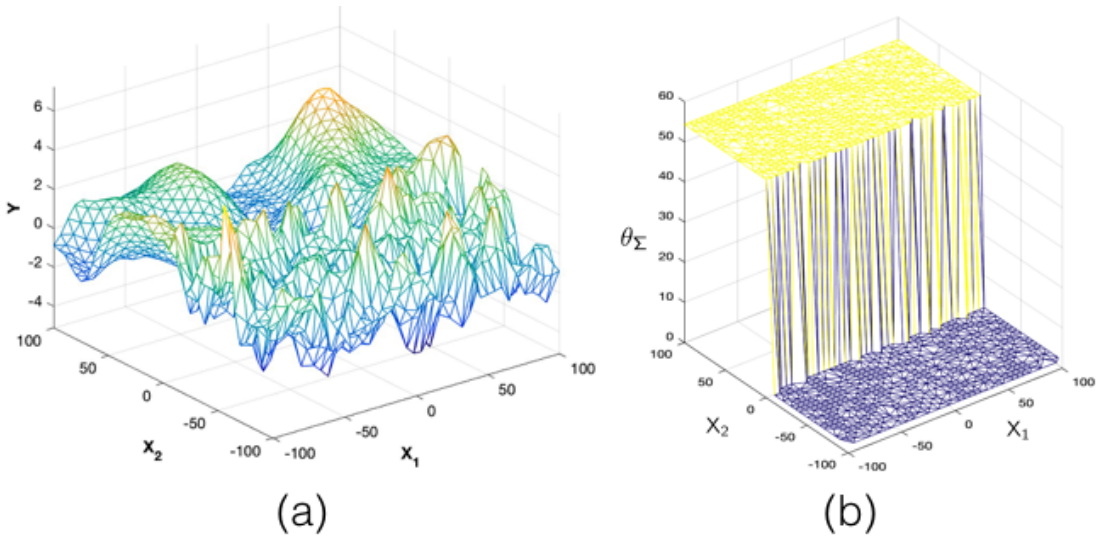


Figure 4.4: Simulation set-up: (a) a sample from the proposed simulation function and (b) a surface from a parameter of the geometric anisotropy (or the scale matrix).

parameters, $\{A, B, C, K, M, \nu\}_w$, of this S-shaped logistic function controls its asymptotic and smoothness properties. See [139] for the various effects of the parameters on the generalized logistic function. Furthermore, the growth rate parameter of B_w controls the smoothness of the simulation function, $y(\cdot)$, across the regions, while the parameter ν_{f_i} of the stationary GP controls the smoothness within the region.

For the geostatistics application, the standard choice of the smoothness parameter in the Matérn function was $\nu_{f_i} = 1$. Moreover, the climate-related processes around the geophysical boundaries, such as land-ocean or plate tectonics, can potentially have a smoothly varying jump; hence, we chose the growth rate parameter of $B_w = 0.5$ for the generalized logistic function. Note, increasing the B_w value also increases the jump in the process parameters around the boundaries. Other values in the generalized logistic function were chosen such that the effects of the GPs were regionalized, i.e., $(A = 0, C = 1, K = 1, \nu = 1, M = 0)$.

For the purpose of our proposed model, in Equation 4.1, we were specifically interested in the spatially-varying parameter of the scale matrix, or the geometric anisotropy. Hence, for the two GPs, $\{f_i\}$, we chose considerably different values for the scale matrix parameters and constant value for the variances. For the signal variance and noise variance, we set $\sigma_f = 1$ and $\sigma_y = 0.1$, respectively. The scale matrix parameters of

anisotropic range and angle, i.e., $\theta_\Sigma = \{\lambda_1, \lambda_2, \psi_A\}_{f_i}$, for the two GPs were set as $(50, 30, 45)_{f_1}$ and $(5, 10, 15)_{f_2}$.[4] To illustrate the variable scale matrix parameters, θ_Σ, Figure 4.4b shows one of the anisotropic range parameters, i.e., $\{\lambda_1\}_{f_i}$, that was selected for the two separate GPs.

To evaluate the performance of various models, we generated 50 realizations of the simulation function with the input space range of $(-100 : 2 : 100)$. Additionally, 50% of the input space samples were uniformly randomly chosen for the training set, $\{\boldsymbol{x}, y\}$, and the rest for the test set, $\{\boldsymbol{x}^*, y^*\}$.

4.4.2 Results

In this section, we evaluate the performance of the Gaussian process model with various covariance functions on the simulation dataset. The covariance functions we compared were the proposed high-dimensional manifold non-stationary covariance function (HD-NSGP, as defined in Equation 4.1), spatially-varying non-stationary covariance function (NSGP, as given in Equation 3.12), and stationary covariance function (SGP, as given in Equation 3.9). Additionally, for the proposed weighted GP simulation function, we ran the known GP parameters of the stationary covariance functions on two separate known regions (TGP). The TGP run was intended to emulate the standard climate scientist approach for modeling the spatial processes based on known geophysical regions (as illustrated in Figure 4.1).

In Table 4.1, we compared the various GP models against the evaluation criteria of MSE, NLPD, and CRPS. The various evaluation metrics are discussed in the Section 4.3.3. The GP models were run for each of the 50 sets of training and test datasets sampled from the proposed simulation function. Moreover, from the experimental setup described in Section 4.4.1, the significance level of 0.01 indicates better performance of the model in Table 4.1.

From Table 4.1, we see that the non-stationary models (NSGP and HD-NSGP) perform better than the stationary model (SGP). For both NSGP and HD-NSGP, the

[4]See Equation 3.16 for the description of the scale matrix parameters.

Table 4.1: Evaluation of various GP models for the proposed simulation function, as given in Equation 4.7. The models compared are the stationary Gaussian process (SGP), spatially varying non-stationary Gaussian process (NSGP), our proposed model of high-dimensional manifold non-stationary Gaussian process (HD-NSGP), and Gaussian process using the known parameters and the covariance function (TGP). The evaluation metrics used are the mean square error (MSE), negative log predictive density (NLPD), and continuous rank probabilistic score (CRPS).

Simulation Function	GP models	SGP		NSGP		HD-NSGP		TGP	
	Eval	Mean	STD	Mean	STD	**Mean**	**STD**	**Mean**	**STD**
Proposed function	MSE	0.33	0.02	0.31	0.01	**0.30**	**0.01**	**0.30**	**0.01**
	NLPD	0.87	0.01	0.80	0.01	**0.78**	**0.01**	**0.77**	**0.01**
	CRPS	0.47	0.2	0.49	0.1	**0.50**	**0.1**	**0.50**	**0.1**

standard deviation (STD) in MSE and NLPD is 0.01, which implies that the models have recovered the true noise variance parameter, i.e., $\sigma_y^2 = 0.01$, of the simulation function. Moreover, the evaluation metrics' means for the HD-NSGP test set runs are better than NSGP, and closest to the TGP model, which suggests that the predictive model of HD-NSGP has learned the true underlying regional GP processes.

For further inspection of the various GP models, we reconstructed the entire surface from the model fit. Figure 4.5 illustrates the surface reconstruction of the GP models.[5] From visual inspection of the surfaces, we see that the non-stationary models have recovered the underlying surface better than the stationary models. The stationary model, as expected, is fairly smooth and homogeneous. Moreover, the HD-NSGP surface has more local areas matched (shown in blue boxes) to the target surface than the NSGP surface. The mismatch (shown in red boxes) in the NSGP surface is primarily at the boundaries. One potential reason for the NSGP mismatch at the boundaries could be that the scale matrix parameters at the boundary locations of the input space are not accurately estimated. Figure 4.6a, illustrates the smooth GP surface of one of the scale matrix parameters in the NSGP model and its inaccurate boundary estimates. In general, a predictive GP model is known to exhibit larger error at the boundaries. In

[5]In Figure 4.5, the input axes are normalized so that the anisotropy can be appropriately bounded in the estimation procedure.

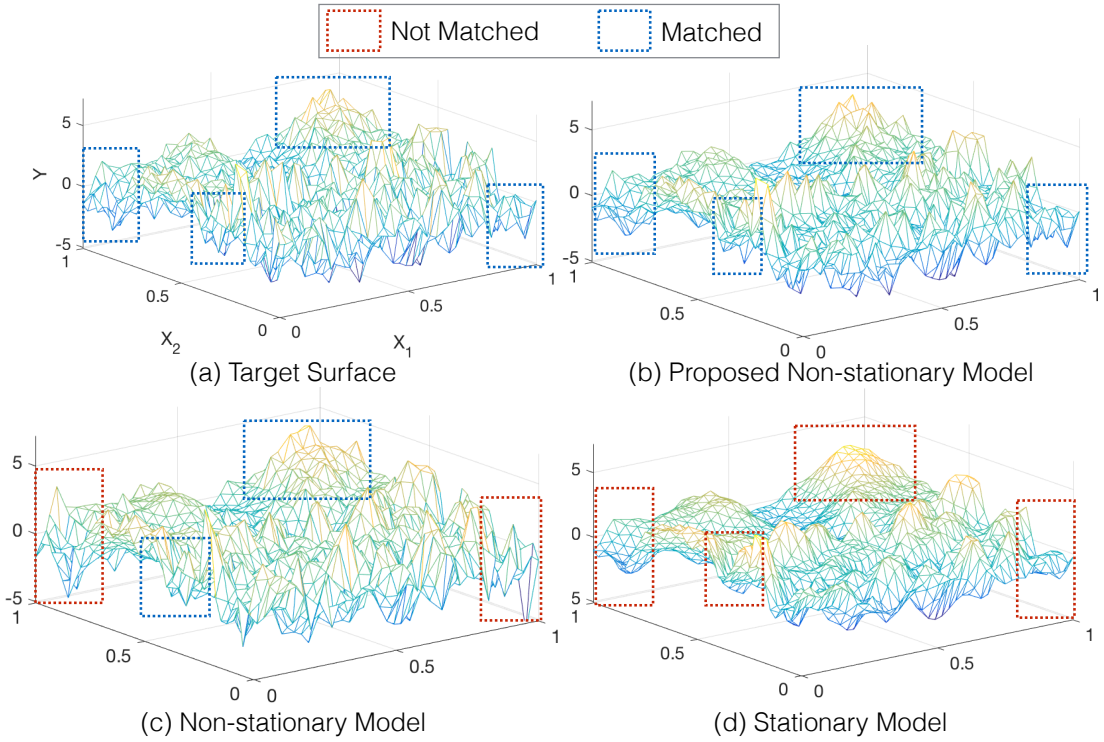


Figure 4.5: Reconstructed surfaces from the Gaussian process (GP) model fit of various covariance functions and the target distribution. The blue boxes show regions matched to the target distribution sample, and the red boxes indicate mismatched regions. The surfaces are (a) generated from a sample of the target distribution, (b) reconstructed from the proposed high-dimensional manifold non-stationary GP model (HD-NSGP), (c) reconstructed from the spatially varying non-stationary GP model (NSGP), and (d) reconstructed from the stationary GP model (SGP).

addition to the predictive GP model for the test dataset, in NSGP there are various GPs that are estimated for the scale matrix parameters. The multiple layers of the GPs in NSGP could potentially increase the prediction errors, especially at the boundaries of the test dataset.

The primary goal of our proposed model (HD-NSGP) is to recover the regionally-varying geometric anisotropy parameter. In Figure 4.6, we depict one of the anisotropic range (for the scale matrix) parameters posterior mean distribution for the non-stationary models. For HD-NSGP, Figure 4.6b shows that it has accurately learned the parameter for the two regional GPs of the simulation function (as illustrated in Figure 4.4b). In comparison, Figure 4.6a shows that the parameter from the NSGP model is oversmooth and inaccurately estimated at the boundaries.

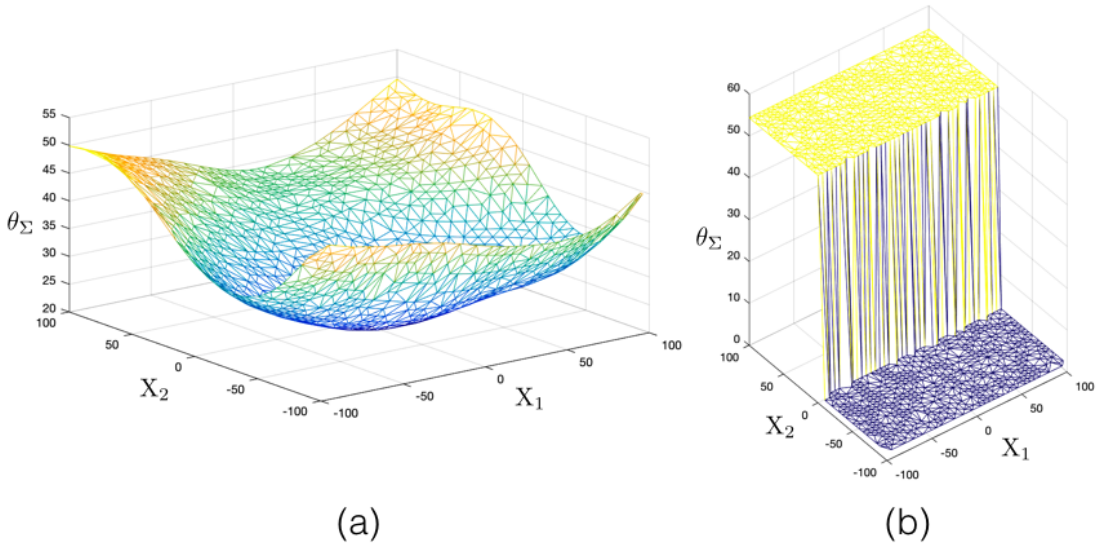


Figure 4.6: Surfaces from one of the learned geometric anisotropy parameters for the (a) spatially-varying non-stationary GP model (NSGP) and (b) our proposed high-dimensional manifold non-stationary GP model (HD-NSGP). The parameter surface from the HD-NSGP model learns the underlying regional geometric anisotropy values of the simulation function.

In Figure 4.6, we show the trace plots of one of the range parameters from the MCMC chains of the non-stationary models. While the burn-in period in the HD-NSGP model (Figure 4.6b) is accomplished after $3 \cdot 10^4$ iterations, the NSGP model (Figure 4.6a) during that period has not yet explored the entire parameter space and is not fully mixed. One reason for the slow mixing and convergence of the NSGP model is that the number of parameters is far more than the HD-NSGP model. Additionally, the various parameters of the GPs in the NSGP's scale matrix model escalates the correlation issues in the parameter estimation procedure. For the two-dimensional input space, the number of parameters required in the NSGP's scale matrix model (as described in Equation 3.19) is $|\theta_{\text{est}}| = 19$, while for HD-NSGP (as described in Equation 4.3) its only $|\theta_{\text{est}}| = 8$ (assuming scalar variance). In higher dimensions, NSGP is infeasible because of the afore mentioned issues in the parameter estimation. In comparison, HD-NSGP shows promise for modeling the non-stationary process in the high-dimensional input space.

The NSGP model in [8] prescribes the Hwang function for its simulation study. In Table 4.2, we compare the performance of various GP models on the samples from the

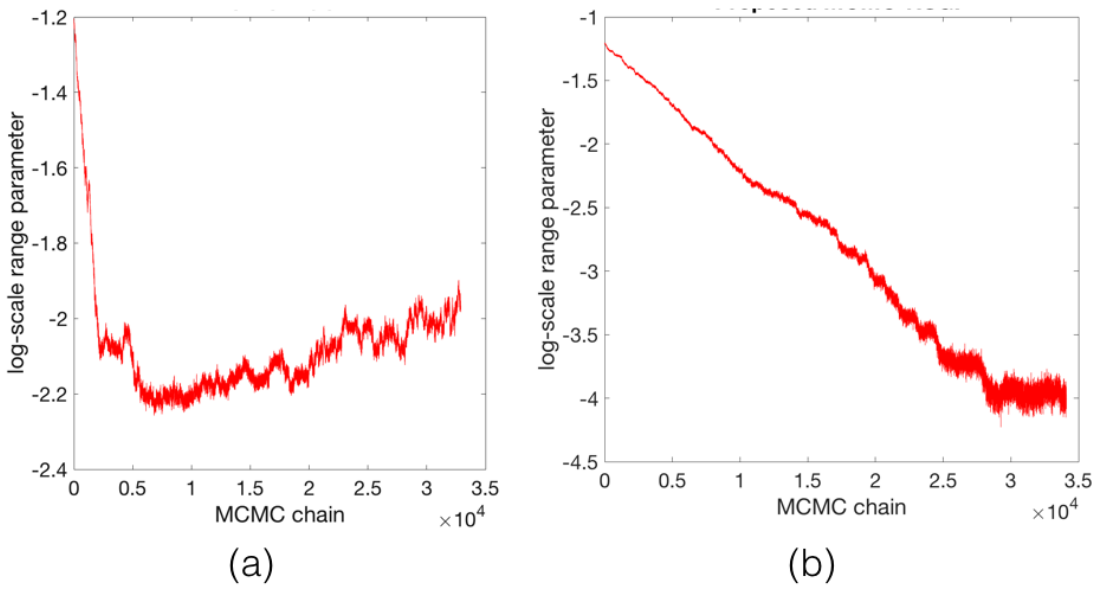


Figure 4.7: Trace plots of one of the geometric anisotropy parameters from the MCMC chains before the burn-in period for the (a) spatially varying non-stationary GP model (NSGP) and (b) our proposed high-dimensional manifold non-stationary GP model (HD-NSGP). The trace plot for the HD-NSGP model shows faster convergence and better exploration of the parameter space when compared to the NSGP model.

Hwang function for the evaluation metric of MSE and NLPD. The GPs were run on all of the 50 sets of the training and test datasets, this setup emulates the scenario of climate-related datasets when the replicates are available. In contrast to the previous experimental setup of the GP run on individual training and test datasets, here we have an increased number of samples to train; hence, a more robust estimate of the large number of non-stationary model parameters can be achieved. In Table 4.2, the significance level of 0.001 indicates better performance of the predictive model. Moreover, here the summary of the GPs' performance comparison is similar to that of the proposed simulation function, i.e., described in Table 4.1. The non-stationary models' performance is better than the stationary models, and HD-NSGP outperforms NSGP, especially in the NLPD evaluation metric. For both of the simulation functions, i.e., Hwang and proposed, the evaluation metrics from Tables 4.1 and 4.2 suggest that HD-NSGP has a better predictive model compared to the SGP and NSGP models. In the next section, we evaluate various GPs on the climate-related datasets that describe sea-level changes.

Table 4.2: Evaluation of the Hwang function: the models compared are the stationary Gaussian process (SGP), spatially varying non-stationary Gaussian process (NSGP), and our proposed model of a high-dimensional manifold non-stationary Gaussian process (HD-NSGP). The evaluation metrics used are the mean square error (MSE) and negative log predictive density (NLPD).

Simulation Function	Eval Metrics	SGP	NSGP	**HD-NSGP**
Hwang's function	MSE	0.0240	0.0237	**0.0235**
	NLPD	0.311	0.278	**0.271**

4.5 Application to Climate Datasets

As stated in the introduction, i.e., Section 4.1, our primary goal for developing the non-stationary model is to improve the estimation of climate-related processes that commonly have a distinct regional variability, such as sea-level changes. For the experiments, we used the following three spatial and spatiotemporal datasets that explains the phenomenon of sea-level changes: glacial isostatic adjustment model output (GIA),[6] tide gauge records (TG),[7] and satellite altimeter observations (SA).[8]

4.5.1 Experimental Setup

The geophysics-driven model output of the glacial isostatics adjustment of vertical land motion (GIA) describes the difference in height between the sea surface and solid earth, as specified in [140]. The GIA model is one of the most prominent geophysical signals present in the sea-level change datasets. The spatial GIA dataset we used is the present-day rate-of-change of relative sea level and crustal uplift prediction in mm/year. Figure 4.8a shows the dense and regularly spaced spatial field of the GIA dataset that we have used. To evaluate the GP's performance on the spatial field, we sampled 25 independent realizations of GIA model output by adding a Gaussian noise of $\sigma_z = 0.2$

[6]The GIA datasets are available at `http://www.psmsl.org/train_and_info/geo_signals/gia/peltier/`

[7]The TG datasets are available at `http://www.psmsl.org/data/obtaining/`

[8]The SA datasets are available at `http://hpc.csiro.au/users/326141/Sea_Level_data/gmsl_files/jb_iby_sry_gtn_giy.nc.gz` or `https://research.csiro.au/slrwavescoast/sea-level/measurements-and-data/sea-level-data/`

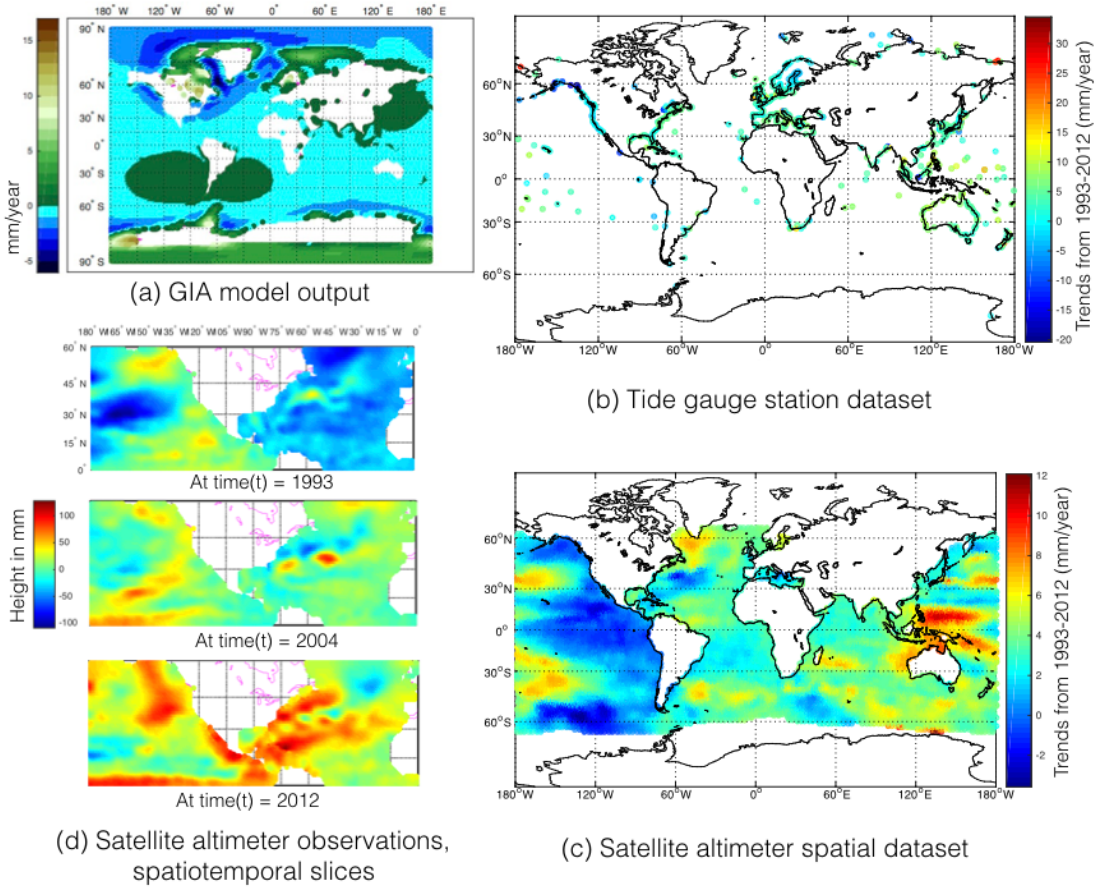


Figure 4.8: Spatial and spatiotemporal climate-related experimental datasets: (a) synthetic dataset (mm/year) of relative sea-level changes from the geophysical model output of the glacial isostatic adjustment model of vertical land motion, (b) trends (mm/year) from 1993 to 2012 of the tide gauge observations, (c) trends (mm/year) from 1993 to 2012 of the satellite altimeter observations, and (d) annual average (mm) of satellite altimeter observations around North America for the time slices 1993 and 2012.

and masking out the land mass. As seen in Figure 4.8a, the response variable is fairly homogeneous and smooth; hence, we used 40% of uniformly random samples from each realization for training and the remainder for testing the GP's performance.

Another climate-related spatial field where we tested the GPs' performance was the rate-of-change in mm/year of the sea-level changes from the monthly tide gauge (TG) station records. The sparse and irregularly spaced monthly tide gauge records from the stations across the globe was prepared by [36]. The time frame we used to estimate the rate-of-change, or trend, was from 1993 to 2012, as there are a maximum number of consistent station records available during this period. Moreover, after 1993 we have

consistent satellite altimeter (SA) observations, which we use in Chapter 5 to align with the TG dataset for the data-fusion experimental setup. The SA dataset measures the ocean surface height from various remote sensing instruments, as detailed in [37]. Figures 4.8b and 4.8c show the trend surface from the TG and SA dataset, respectively. Because the tide gauge stations are sparsely spaced, we used 30% of uniformly random region-wise hold-out samples to evaluate the GPs' predictive performance and the remainder to train the model. Moreover, we prepared regional training and test sets, such as North America's geophysical regions as illustrated in Figure 4.1b, to evaluate the GPs.

When including the spatially dense datasets of GIA and SA in the experiments, the input space resolution was subsampled by taking the mean of the values within a 5° grid as the response variable. For all of the spatial datasets, the response variable (trend) was spatially demeaned and normalized so that we could evaluate primarily the covariance structure across the various datasets.

Finally, we used the spatiotemporal SA dataset to evaluate our proposed GP model in high dimensions. Figure 4.1d shows the time slice of the dense and regularly spaced SA dataset from regions around North America. To focus our evaluation on the covariance structure of the spatiotemporal dataset, we modeled the forced signals in the sea-level changes using the following formulation:

$$y(\boldsymbol{s}, \boldsymbol{t}) = f(\boldsymbol{s}, \boldsymbol{t}) - g(\boldsymbol{t}) - l(\boldsymbol{s})(\boldsymbol{t} - \boldsymbol{t_0}). \tag{4.8}$$

The above principal formulation has been used in the sea-level community to explain the forced signals that occur in regional relative sea-level changes (see [26]). Here, the response variable, $y(\cdot, \cdot)$, represents the emergence of the forced signals at the geolocations $\boldsymbol{s}$ and time $\boldsymbol{t}$, i.e., $\boldsymbol{x} = \{\boldsymbol{s}, \boldsymbol{t}\}$; the sea level measurements from the dataset are $f(\cdot, \cdot)$; $g(\cdot)$ is the global mean sea level (GMSL); and $l(\cdot)$ is the local mean sea level trend (LMSL) at each geolocation.[9] To compute GMSL, we took the spatial mean

[9]The term LMSL when referred to the TG dataset is also commonly called as the relative sea-level changes (RSL). RSL is illustrated in Figure 1.2a.

at each time slice of the dataset, and for the LMSL computation we used the trend surface from the aforementioned spatial experimental setup. For our spatiotemporal experimental setup, we restricted the spatial range around the North American region and the temporal range from 1993 to 2012. The time series was annually averaged and demeaned for the SA dataset. The hold-out set of 30% of uniformly random samples were used to evaluate the GPs' performance and the remainder to train the model.

We report the performance of various GP models with respect to two evaluation metrics, i.e., the mean squared error (MSE) and negative log predictive density (NLPD).

4.5.2 Results

In Table 4.3, we compare the various GP models for the aforementioned spatial and spatiotemporal climate-related experimental setup. From Tables 4.1, 4.2, and 4.3, we see a consistent result where the non-stationary models performs better than the stationary models, and our proposed non-stationary model (HD-NSGP) improves upon the NSGP model.

For the GIA dataset, the difference between the non-stationary models' evaluation metrics is small; this could be because of the homogenous and dense spatial field. In other words, there is little variation within the regions, and the boundaries are smooth (as seen from Figure 4.8a). In Figure 4.9, we illustrate the mean posteriors of the scale matrix parameter at the geospatial boundaries around the regions of the Barents sea from the GIA dataset. The scale matrix parameter in the spatial dataset is represented using an ellipse, and the relative size of the ellipses are scaled for illustration. While the scale matrix parameter from the SGP model (illustrated in Figure 4.9b) has the same shape for all of the input space locations, NSGP and HD-NSGP models show the spatially-varying shape of the scale matrix parameter. Moreover, the shape of the ellipses from the NSGP model, at different regions, has less variability, or is over-smoothed, compared to that of the HD-NSGP model. This over-smoothing in the NSGP model could potentially be the reason for its slight underperformance in Table 4.3 when compared to the HD-NSGP model. For all of the non-stationary models, the rotation parameter of the scale matrix shows an appropriate fit to the underlying field. This

Table 4.3: Evaluation of various GPs for the spatial and spatiotemporal climate-related datasets at global and regional extents. The spatial datasets used are the Glacial-isostatic Adjustment (GIA) and tide gauge records (TG), and the results are in the units of mm/year. The spatiotemporal dataset used is the satellite altimeter observations (SA), and the results are in the units of mm. The mean of posteriors from the GP models of SGP, NSGP, and HD-NSGP, are compared using the evaluation metrics of MSE and NLPD.

Dimensions	Datasets	SGP		NSGP		HD-NSGP	
		MSE	NLPD	MSE	NLPD	**MSE**	**NLPD**
Spatial ($m = 2$)	GIA (Global)	0.58	3.08	0.56	2.00	**0.54**	**1.94**
	GIA (Barents Sea)	1.57	9.23	1.24	7.30	**1.03**	**6.90**
	TG (Global)	0.85	2.81	0.75	2.80	**0.71**	**2.78**
	TG (N.America)	1.05	3.0	0.54	2.62	**0.50**	**2.58**
Spatiotemporal ($m = 3$)	SA (N.America)	1.38	4.29	1.18	4.25	**1.10**	**4.15**

fit of the ellipse is observed from its minor axis being approximately rotated along the direction of the most prominent gradient at a particular location in the field.

Figure 4.10 illustrates the estimates of the scale matrix parameters for the spatiotemporal SA dataset at a particular instance, i.e., at $t = 2004$. Here, the shape of the ellipses corresponds to the spatial scales, while the colors represent the temporal scale. Moreover, the figure emphasizes the difference in the estimates of the spatiotemporally constant (Figure 4.10a) and varying (Figure 4.10b) scale matrix parameters from SGP and HD-NGP models, respectively. This difference in the stationary and non-stationary covariance structure in high-dimension improves the predictive model, as seen from Table 4.3.

Similar to the spatial GIA dataset, the spatiotemporal SA dataset is dense, regularly spaced, and spatially subsampled for our prototypical climate-related experimental setup. Furthermore, the temporal scale is annually averaged, and the temporal span is short, i.e., only 19 years. This experimental setup could be one of the reasons that we

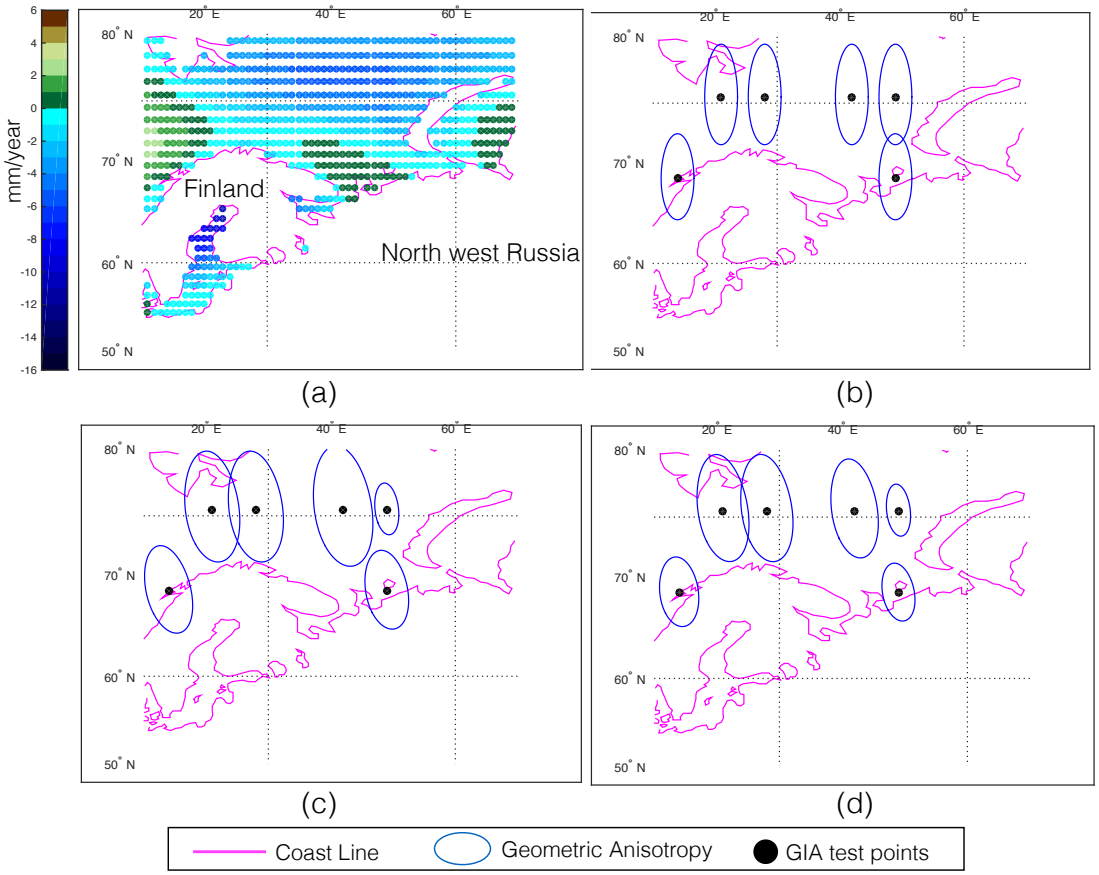


Figure 4.9: A regional analysis of the glacial-isostatic adjustment (GIA) dataset. (a) GIA dataset around the Barents sea region. The mean posteriors of the scale matrix, or geometric anisotropy, parameters from the (b) stationary GP model (SGP), (c) spatially varying non-stationary GP model (NSGP), and (d) our proposed high-dimensional manifold non-stationary GP model (HD-NSGP). Compared to the SGP model, the NSGP and HD-NSGP models shows a better conformity of the scale matrix parameter to the underlying spatial field.

were unable to find a distinct spatiotemporal regional variation from our non-stationary model fit. In other words, we found that the estimated scale matrix shapes were overall similar from both the NSGP and HD-NSGP models. Even with a slight difference in the estimates of the spatiotemporally-varying scale matrix parameters between the NSGP and HD-NSGP model, we see an improvement in the predictive performance from Table 4.3. This intuitively exemplifies that the spatially-varying non-stationary GP models predictive performance is sensitive to the estimates of the scale matrix parameter for complex climate-related processes.

A more distinct regional variation is seen in the TG dataset when compared to the

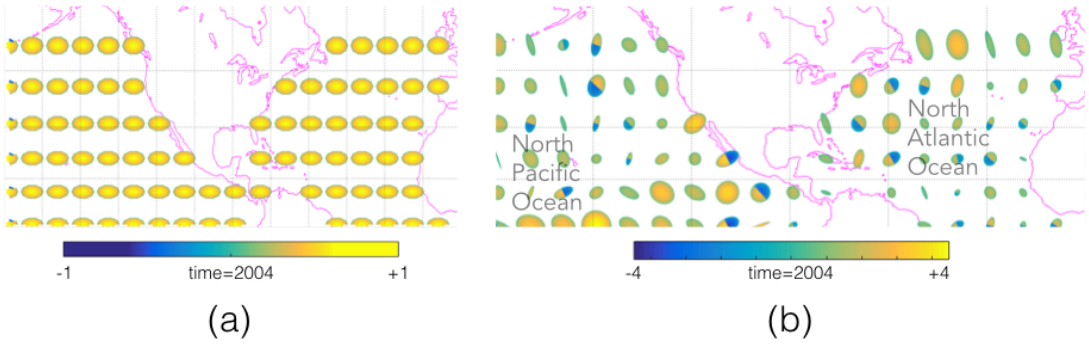


Figure 4.10: An estimate of the scale matrix parameter, or geometric anisotropy, for the spatiotemporal satellite altimeter dataset. The mean posteriors of the scale matrix parameter shown here are from the (a) stationary GP model (SGP) and (b) our proposed high-dimensional manifold non-stationary GP model (HD-NSGP).

GIA and SA datasets. For example, Figure 4.11a shows color-coded clusters of stations around North Americas that are advised by the geophysics of the regions. Moreover, the estimates, or the mean posteriors, of the scale matrix parameter from our proposed model (HD-NSGP), NSGP model, and SGP model are represented by the ellipses in Figures 4.11a, b, and c, respectively. These estimates of ellipses from various GP models show a noticeable difference in its patterns.

While the NSGP model estimates the scale matrix parameters that are spatially varying, it fails to capture the underlying regional variability of the coastal stations. One of the potential reasons that the NSGP model gives a large variability of the scale matrix estimates across the nearby stations is because there are insufficient spatial neighbors in the input space. Similar behavior of the scale matrix estimates from the NSGP model was seen in our simulation study (see Figure 4.6). On the other hand, our proposed model (HD-NSGP) aligns with the orientation of the nearby station data at the coastline. Also, the clusters represented by the boxes in Figure 4.11b conform with the geophysical clusters. Moreover, Table 4.3 shows improved performance of HD-NSGP model for North America's TG test dataset. Hence, our proposed HD-NSGP model is particularly helpful in cases where there is distinct regional variability in the underlying physical processes.

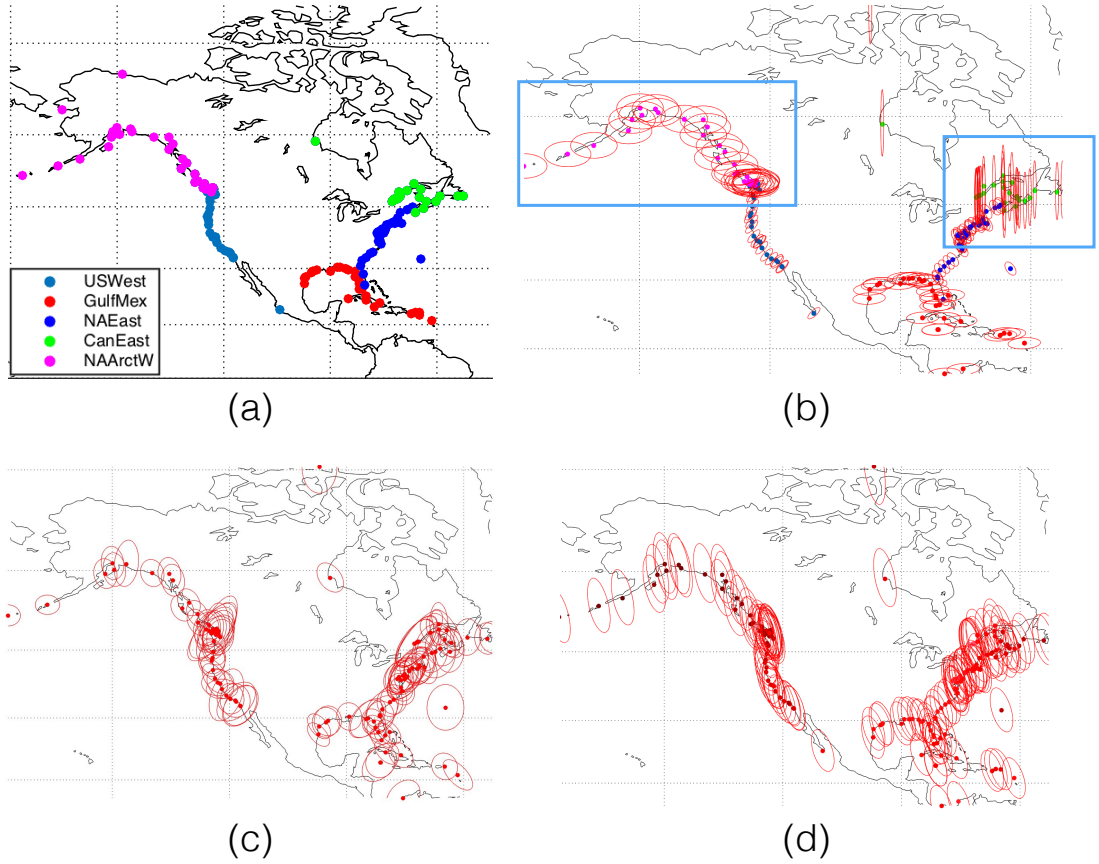


Figure 4.11: A regional analysis of the tide gauge station records (TG) around North America. (a) Spatial clusters of stations (i.e., same color) advised by the geophysics of the region. The mean posteriors of the scale matrix, or geometric anisotropy, estimated from (b) our proposed high-dimensional manifold non-stationary Gaussian process (HD-NSGP) model, (c) spatially-varying non-stationary (NSGP) model, and (d) stationary (SGP) model. The clusters of the estimated scale matrix parameters from HD-NSGP, as shown in the blue boxes in (b), conforms with the geophysical clusters, as shown in (a).

4.6 Discussion

In this chapter, we proposed a high-dimensional manifold non-stationary covariance function (HD-NSGP) and its parameter estimation scheme for the GP model. Furthermore, in our experiments, we showed that HD-NSGP outperformed both the previously developed non-stationary GP in [8] (NSGP) and the stationary GP (SGP). The datasets considered in our experiments were from the simulation study, climate model output (i.e., GIA model), remote sensing dataset (i.e., satellite altimeter (SA)), and station data (i.e., tide gauge (TG) stations).

For the TG dataset, we showed that HD-NSGP successfully modeled the covariance structure of the underlying distinct regional variability around North America. For the smooth, dense, and regularly spaced datasets of GIA and SA, the scale matrix process of HD-NSGP and NSGP showed similar outcomes. Resembling the TG dataset, the climate-variable of precipitation also exhibits distinct regional characteristics in its spatial process. For example, in Figure 1.1 we depict the trend surface of precipitation around North America and its potential regional component. Future work can focus on exploring HD-NSGP's applicability to various climate-related processes.

A crucial aspect of developing a non-stationary model for the complex climate-related processes is the simulation study. In Section 4.4, we proposed a flexible framework for testing a non-stationary model, and, further showed the applicability of our proposed model. A straightforward extension of our simulation function, in Equation 4.7, could include multiple regional GPs with differing smoothness parameters in the covariance and generalized logistic functions.

For the methodological development in HD-NSGP, we modeled the spatially-varying scale matrix, or geometric anisotropy, on a Riemannian manifold of SPD matrices. This manifold construction provided us with a valid covariance function and the essential smoothness properties of its associated GP. In comparison, NSGP modeled each element of the eigendecomposed scale matrix as a separate GP, resulting in various issues in the NSGP model, as described in Section 3.2.3. For example, the scale matrix parameters in NSGP is $|\theta_\Sigma| = 16$ (see Equation 3.19), while in HD-NSGP it is only $|\theta_\Sigma| = 8$ (see Equation 4.3). Furthermore, the computational complexity for estimating the scale matrix in NSGP is $\mathcal{O}(m^2 \cdot n^3)$. In comparison, for the simplified HD-NSGP model it is $\mathcal{O}(m^6)$, and for the full model it is $\mathcal{O}(|r| \cdot m^6)$ (as described in Equation 4.2). Note, in many of the spatial problems, including the applications described in this chapter, $m << n$ and $|r| < n$. Hence, by reducing the number of parameters in HD-NSGP, we render it scalable and suitable for a high-dimensional input or covariate space.

Moreover, we sample the scale matrix using the intrinsic statistics on a manifold, leading to its uniqueness. In other words, each element of the scale matrix is represented as a coordinate system with its canonical intrinsic metric. Hence, HD-NSGP

does not encounter the identifiability issues in its scale matrix estimation, which is inherent in the NSGP model. In addition, HD-NSGP's scale matrix model is flexible to vary its smoothness within and across the regions, as illustrated in Section 4.4.2. The spatially-varying smoothness of the scale matrix process is a result of computing the neighborhood structure of the scale matrix on a smooth manifold, as described in Section 4.3.2. The neighborhood structure of the scale matrix can be further explored using recent advances in the learning techniques on a Riemmanian manifold [141].

In our simulation study, we saw that the NSGP's scale matrix process tends towards an over smooth estimate. One possible extension to overcoming this smoothness issue could include using the Matérn covariance function instead of the squared exponential function used in the NSGP model of [8]. Also, in theory, one could increase the depth of the NSGP model hierarchy by using spatially-varying covariance functions for the scale matrix model's GPs. However, in practice, we found that both of the strategies mentioned above for incorporating varying smoothness in the NSGP scale matrix model increased its existing parameter estimation issues, such as identifiability, convergence, and mixing.

In Figure 4.7, we illustrated the mixing and convergence improvement of HD-NSGP when compared to NSGP. The mixing and convergence issues of NSGP is described in detail in Section 3.2.3. One of the reasons MCMC samples mix well for HD-NSGP is because we jointly sample the scale matrix. Similarly, MCMC converges faster in HD-NSGP when compared to NSGP for reasons including a reduced number of matrix inversions, hyperparameter space, and parameter space for the scale matrix process, as described in Section 4.3.2.

Albeit, our goal for developing the parameter estimation scheme was to primarily show HD-NSGP's applicability to climate-related processes, such as sea-level change. Hence, we tailored the two intractable, or problematic, parameters informed by the climate-related processes, i.e., the smoothness parameter of the Matérn function (ν) and the number of regions, or clusters, ($|r|$). For example, it is well known in spatial statistics (see [1]) that many climate-related processes have the smoothness parameter that approximately lies in the range of $\nu = \{0.5, 1, \ldots, 2.5\}$. Similarly, the range of

the number of clusters were approximately bounded by the geophysical boundaries of the coastline or by visual inspection. For example, in North America's TG dataset the range of the clusters were $|r| = \{2, \ldots, 10\}$, and informed by the coastline. While in the SA dataset the range of the clusters, $|r| = \{n/2, \ldots, n\}$, was chosen by visual inspection. A potential extension in HD-NSGP's general applicability to other stochastic processes where the underlying process is non-stationary could benefit from a more robust estimation procedure. One such direction is to develop a local-likelihood based estimation procedure similar to [113].

In addition, there are two areas of our covariance model that can be further improved. First, is the intrinsic statistics of the SPD matrices, and second is the Matérn covariance function of the GP. The aspects of the manifold that could provide insights into the scale matrix model development are the choice of metric, connections, and the vector space. We elaborate on potential improvements of the manifold representation in Chapter 7, as its the common theme in our approach to the geostatistical problems of Chapters 4, 5, and 6.

To improve the spatially-varying non-stationary covariance function, we have shown how the high-dimensional manifold representation of one of the covariance function's parameters, i.e., the geometric anisotropy, can be modeled. Other parameters in the covariance function, such as the signal variance and noise variance, could similarly be modeled as the spatially-varying anisotropy on the manifold of SPD matrices. Moreover, the spatially-varying smoothness parameter in the Matérn, as described in Equation 3.14 and derived in [18], can be implemented in conjunction with our HD-NSGP model.

Another aspect of the spatial statistics that is actively being developed in recent years is its application to large datasets. Methods that have shown promise for this task include the Lattice Kriging [96], Nearest Neighbor GP [97], and stochastic PDE [118]. In general, the aforementioned methods for large datasets leverage the sparse representation of the precision matrix instead of modeling for the covariance matrix. An interesting future development of our approach could involve the manifold representation to solve large, high-dimensional, non-stationary datasets that are commonly found

in climate-related problems. On the other hand, there are many climate-related datasets that still suffer from a small sample size, such as TG dataset. In the next chapter, we provide extensions to our proposed non-stationary model for improving inference from small datasets using a data-fusion technique.

Chapter 5

Data-Fusion Geostatistical Model

Data, data, data, so many sources to consider, yet so few . . . can we fuse them all to understand the phenomena better?

5.1 Introduction

Any predictive model is only as good as the estimated parameters that represent the data. Many climate applications, including sea-level change predictions, lack the availability of recorded measurements for robustly estimating the parameter. For example, Figure 5.1a shows the lack of temporal range data available in the satellite altimeter (SA) dataset. On the other hand, Figure 5.1c shows the spatially dense availability of spatiotemporal SA dataset. In contrast, Figure 5.1a shows the temporally long range of data available at one of the tide gauge (TG) sites, and Figure 5.1c shows the spatially sparse availability of spatiotemporal TG dataset.

In Section 1.3, we briefly described the problem of inference from multiple sources of datasets. In climate-related applications, such as the time series of global mean sea-level change, [27] proposed a standard principal component analysis (PCA) based method for inference from the TG and SA datasets. Various other regional and climate-variable specific inference methods from multiple sources that primarily rely on the geophysics can be found in [20].

In the climate and statistics literature, a systematic data-driven approach for multi-source inference is still ongoing research. Furthermore, to the best of our knowledge, we have not found a non-stationary model that also incorporates multiple sources of datasets for modeling climate-related variables, such as sea-level changes.

Hence, in this chapter, we propose an extension to our high-dimensional manifold

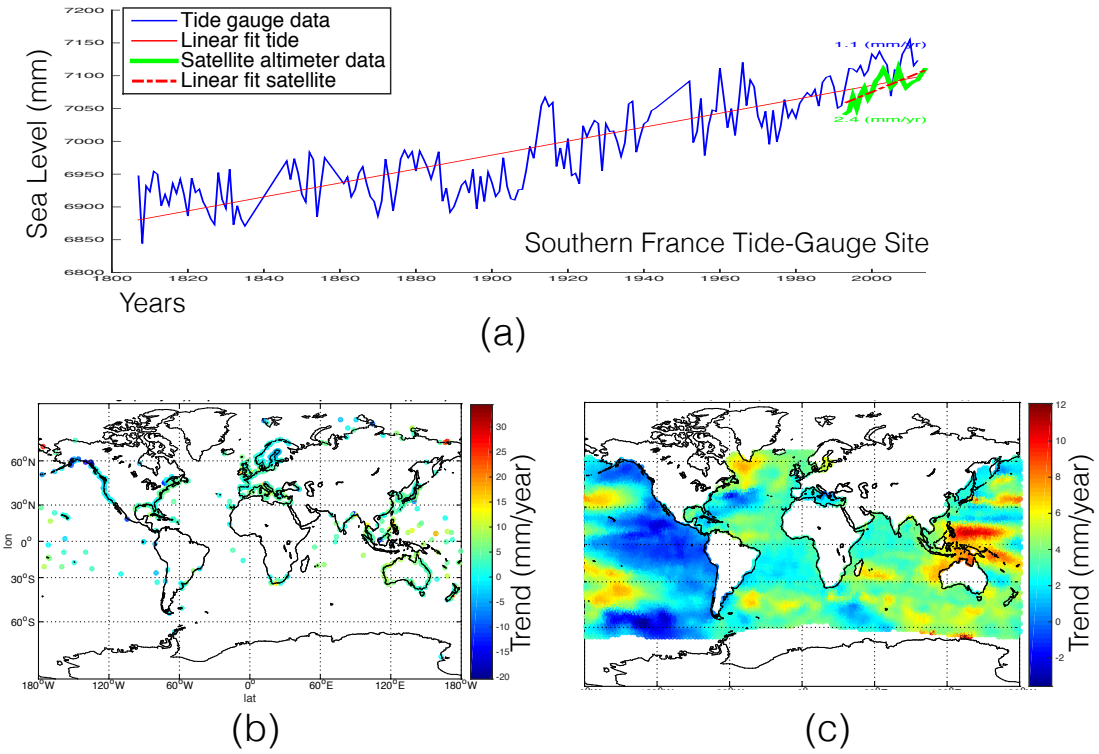


Figure 5.1: Spatiotemporal range in the various sources of sea-level datasets. (a) An example of the temporal range at the oldest tide-gauge site that is located in southern France. An example of the available spatial measurements around the globe from (b) tide-gauge sites and (c) satellite-altimeter dataset. Note, the satellite-altimeter dataset is spatially dense, and the tide-gauge dataset is temporally long ranged.

non-stationary Gaussian process (HD-NSGP) model of Chapter 4, which incorporates multiple sources of datasets for improving inference. The core idea is to *fuse* the estimates of the parameters, such as the spatially-varying scale matrices, from multiple sources of datasets using the manifold construction of Chapter 2. We call this framework a data-fused high-dimensional manifold non-stationary Gaussian process (DF-NSGP) and describe this model in the next section.

5.2 Data-Fused High-dimensional Manifold Model

In Figure 5.2, we depict the DF-NSGP model, where the spatially-varying parameter of the geometric anisotropy, or scale matrix, from the two sources of sea-level measurements are estimated on a manifold, $\mathcal{M}$, of SPD matrices. The proposed correlation function, dissimilarity function, and spatially-varying geometric anisotropy for

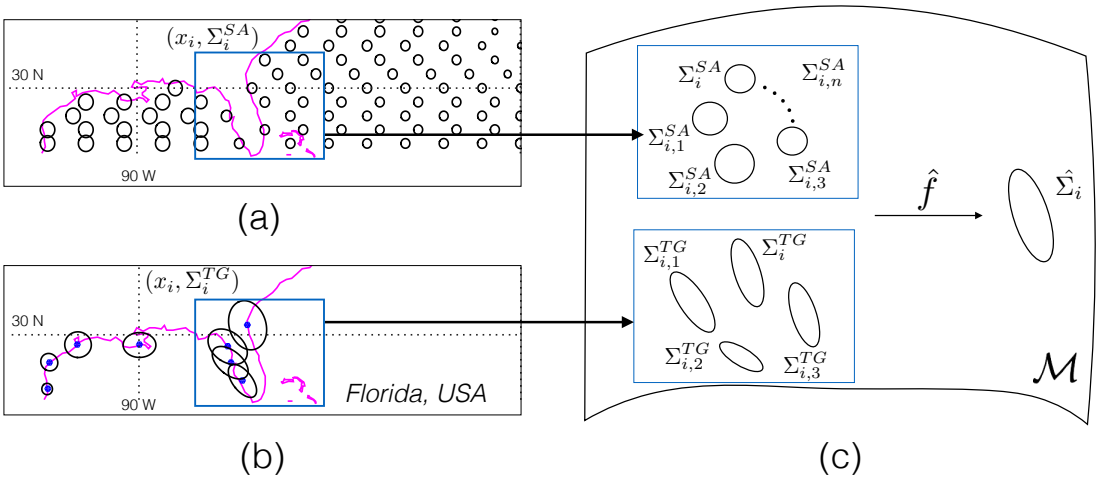


Figure 5.2: The proposed data-fusion framework: (a) tide-gauge sites (b) satellite-altimeter dataset, and (c) data fusion of the geometric anisotropy on a manifold.

DF-NSGP is constructed using the following equations:

$$\rho^{DFNS}(\boldsymbol{x}_i, \boldsymbol{x}_j) = 2^{\frac{m}{2}} |\hat{\Sigma}_i|^{\frac{1}{4}} |\hat{\Sigma}_j|^{\frac{1}{4}} \left|\frac{\hat{\Sigma}_i + \hat{\Sigma}_j}{2}\right|^{-\frac{1}{2}} \rho^S(\nu, q_{ij}^{DFNS}), \tag{5.1a}$$

$$q_{ij}^{DFNS} = (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \left(\frac{\hat{\Sigma}_i + \hat{\Sigma}_j}{2}\right)^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j), \tag{5.1b}$$

$$\hat{\Sigma}_i \sim f(\hat{\Sigma}_{r(i)}^{TG}, \hat{\Sigma}_{r(i)}^{SA}), \quad \hat{\Sigma}_j \sim f(\hat{\Sigma}_{r(j)}^{TG}, \hat{\Sigma}_{r(j)}^{SA}), \tag{5.1c}$$

$$\hat{\Sigma}_{r(i)}^{TG} \sim \mathcal{N}(\bar{\Sigma}^{TG}, \Lambda^{TG}), \quad \hat{\Sigma}_{r(i)}^{SA} \sim \mathcal{N}(\bar{\Sigma}^{SA}, \Lambda^{SA}). \tag{5.1d}$$

Note, Equations 5.1a, 5.1b, and 5.1d are similar to those of the HD-NSGP model, as given in Equation 4.1. The only difference is the addition of Equation 5.1c, which samples $\hat{\Sigma}_i$ using the estimation function, $f(\cdot)$, and the spatially-varying scale matrices from the two sources of datasets. The estimates of the scale matrices from the two sources of datasets are denoted as Σ^{TG} and Σ^{SA}, referring to the tide gauge and satellite altimeter datasets, respectively. In general, we denote the primary source of the scale matrix estimate as Σ and the surrogate source as Σ^{Surr}.

A straightforward estimation, or regression function, $\hat{\Sigma}_i = f(\Sigma_{r(i)}^{TG}, \Sigma_{r(i)}^{SA})$, can be

represented as follows:

$$\hat{\Sigma}_i = \boldsymbol{\phi_1} \odot \Sigma_{r(i)}^{TG} \oplus \boldsymbol{\phi_2} \odot \Sigma_{r(i)}^{SA} \oplus \boldsymbol{\phi_3}, \tag{5.2}$$

where the hyperparameters $\{\boldsymbol{\phi_i}\} = \{\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \boldsymbol{\phi_3}\}$ transform the initial estimates of the scale matrix, $\{\Sigma_{r(i)}^{TG}, \Sigma_{r(i)}^{SA}\}$, which are estimated from the individual sources of datasets. For a matrix-valued hyperparameter, $\boldsymbol{\phi_i} \in GL(m, \mathbb{R})$, the tensor operations, $\{\odot, \oplus\}$, on the tangent space of a manifold, $\mathcal{T}_\Sigma\mathcal{M}$, endowed with the affine-invariant metric can be defined as follows:

$$\begin{aligned} \boldsymbol{\phi_i} \odot \Sigma &= \exp_\Sigma(\log_\Sigma(\boldsymbol{\phi_i}) \cdot \Sigma), \\ \Sigma \oplus \Sigma' &= \exp_\Sigma(\Sigma + \log_\Sigma(\Sigma')), \end{aligned} \tag{5.3}$$

where the operations $\{\cdot, +\}$ are the usual matrix multiplication and addition operation on $\mathrm{gl}(m, \mathbb{R})$. The definition of the exponential and logarithm maps are described in Section 2.2.2. Furthermore, the above formulation can be computed using the spectral decomposition of the SPD matrices.

The above operation of $\{\oplus\}$ uses the parallel transport construction because Σ and Σ' can possibly lie on separate tangent spaces, i.e., $\mathcal{T}_\Sigma\mathcal{M}$ and $\mathcal{T}_{\Sigma'}\mathcal{M}$, respectively. However, we still assume that there exists a curve that passes through the two SPD matrices and that they lie on a locally affine diffeomorphic neighborhood. See Appendix A.3 for the properties of the geodesic distance function on a statistical manifold. When the SPD matrices are assumed to lie on the same tangent space, the usual addition operation, $\{+\}$, for matrices can be applied. Given that $\boldsymbol{\phi_i} \cdot \Sigma \notin \mathrm{S}^+(m, \mathbb{R})$, the usual matrix multiplication operation, $\{\cdot\}$, cannot be used. Alternatively, a computationally favorable logarithmic multiplication operation exists using the bi-invariant metric structure on a manifold,[1] as given in [10], and is defined as follows:

$$\begin{aligned} &\boldsymbol{\phi_i} \odot \Sigma = \exp(\log(\boldsymbol{\phi_i}) + \log(\Sigma)) \quad \text{when } \boldsymbol{\phi_i} \in \mathrm{GL}(m, \mathbb{R}) \quad \text{and} \\ &\phi_i \odot \Sigma = \exp(\phi_i \log(\Sigma)) \quad \text{when } \phi_i \in \mathrm{GL}(1, \mathbb{R}). \end{aligned} \tag{5.4}$$

[1]The logarithmic bi-invariant metric structure on a Riemannian manifold is also called the Log-Euclidean metric structure.

For simplicity, in our experiments we used scalar hyperparameter values of $\phi_i \in \mathrm{GL}(1, \mathbb{R})$. However, a computationally involved and favorable choice would be to either implement a matrix-valued hyperparameter, $\boldsymbol{\phi_i} \in \mathrm{GL}(m, \mathbb{R})$, or a spatially-varying hyperparameter, $\boldsymbol{\phi_i}(\boldsymbol{x})$. Alternatively, instead of the regression function in Equation 5.2, we can assume that Σ^{TG} and Σ^{SA} are sampled from the same distribution on, $\mathcal{N}(\bar{\Sigma}, \Lambda)$, the SPD manifold. In other words, the data-fused scale matrix is sampled from $\hat{\Sigma}_i \sim \mathcal{N}(\bar{\Sigma}, \Lambda \mid \Sigma^{TG}_{r(i)}, \Sigma^{SA}_{r(i)})$, and the hyperparameters are $\{\phi_1 = 0.5,\ \phi_2 = 0.5,\ \phi_3 = 0\}$. For our experiments, we used this formulation of the data-fused scale matrix since it simplifies the computational complexity in the sampling procedure. Another alternative for estimating the hyperparameters, $\{\phi_i\}$, is to use the geophysical relation between the different sources of measurements. For example, in our experiments, we can separately learn the geophysics-driven geometric anisotropy relationship between the tide gauge records and the satellite altimeter measurements, and then fix the hyperparamter values in Equation 5.2. Furthermore, below we state the relevant theorems of a data-fused non-stationary covariance function and its smoothness properties.

Theorem 3. *The spatially-varying non-stationary covariance function, as constructed from Equation 5.1, is a valid non-stationary covariance function.*

Proof. The proof for the above statement follows the same recipe as described in Theorem 1. That is, by construction, on a smooth manifold with the affine-invariant metric we are guaranteed an outcome of the scale matrix estimation procedure to be an unique SPD matrix, i.e., $\hat{\Sigma}_i \in \mathrm{S}^{+}(m, \mathbb{R})$. Hence, our prescription of the non-stationary covariance function, as constructed from Equation 5.1, is valid. □

Theorem 4. *A non-stationary Gaussian process, $Z(\cdot)$, is Mth-order sample path and mean square differentiable for the non-stationary correlation function, $\rho^{DFNS}(\cdot, \cdot)$, as given in Equation 5.1, if the associated stationary correlation function, $\rho^{S}(\cdot)$, is Mth-order sample path and mean square differentiable.*

Proof. The proof of the above statement follows the same recipe of Theorem 2. The data-fused scale matrix, $\hat{\Sigma}_i$, is sampled from a distribution that lies on an infinitely

differentiable and complete Riemannian metric space. Additionally, we assume without loss of generality that the hyperparameters that provide the relation between the two sources of datasets, i.e., $\{\boldsymbol{\phi_i}(\boldsymbol{x})\}$, are either constant or infinitely differentiable. Hence, by construction, the smoothness properties of the sample path and mean square differentiability for the non-stationary Gaussian process only depends on the associated, chosen stationary correlation function. □

Proposal Scheme 4: For the spatially-varying geometric anisotropy, or scale matrix parameters, $\{\hat{\Sigma}_i\}$, of DF-NSGP model.

1: Initialize the samples, $\theta_\Sigma = \{|r|, \bar{\Sigma}, \Lambda\}$, using Proposal Scheme 2.

2: Use the pre-estimated mean posterior samples of the parameters from the surrogate source of the dataset, i.e., $\{\Sigma^{\text{Surr}}_{r(i)}\}$, using the HD-NSGP model. The nearest neighbor region in the surrogate source can be established by using nearest neighbor geolocation in the two sources of datasets, i.e., $\boldsymbol{x_j}^{\text{Surr}} := \text{argmin}(\{d_{gcd}(\boldsymbol{x_i}, \boldsymbol{x_j}^{\text{Surr}})\}_{j=1,\ldots,n^{\text{Surr}}})$.

3: If using Equation 5.3, sample $\boldsymbol{\phi_i} \sim \mathcal{N}(\bar{\Sigma}_{\boldsymbol{\phi}}, \Lambda_{\boldsymbol{\phi}})$ for $\boldsymbol{\phi_i} \in \text{S}^+(m, \mathbb{R})$ using Proposal Scheme 2 and Steps 2 to 5.

4: Else, sample $\hat{\Sigma}_i \sim \mathcal{N}(\bar{\Sigma}, \Lambda | \Sigma_{r(i)}, \Sigma^{\text{Surr}}_{r(i)})$ such that $\text{tr}(\Lambda) \leq \tau^2$, where τ is either a constant threshold or is sampled from $\tau \sim \mathcal{N}(0, 1)$.

For estimating the HD-NSGP parameters, we used a similar MCMC scheme, as described in Section 4.3.2. The only difference is in the estimation of the geometric anisotropy, or scale matrix. Instead of using Proposal Scheme 3 for HD-NSGP, we used Proposal Scheme 4 for DF-NSGP. For simplicity, we used the pre-estimated mean posteriors of the scale matrix parameter, Σ^{Surr}, from the surrogate source. The threshold in Step 4 of Scheme 4 was used to restrict the dispersion in the scale matrix parameters from the two sources of datasets. Additionally, this threshold also acted to preserve the assumption of using data fusion. In other words, we used the data-fused sampler only if the associated region of the surrogate scale matrix parameter was in the local vicinity on the manifold. Interesting future work would be to simultaneously sample

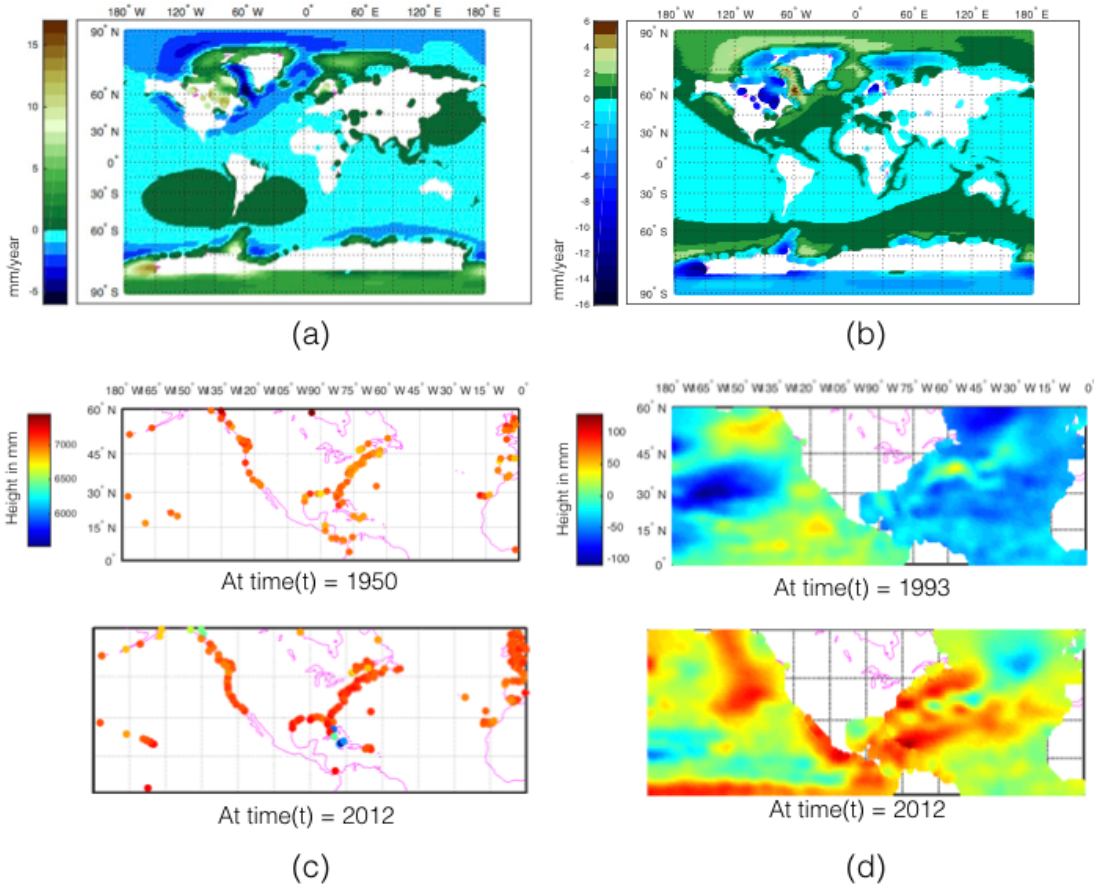


Figure 5.3: Data-fusion experimental setup: (a) spatial glacial isostatic adjustment of vertical land motion (GIA-VLM) in mm/year, (b) spatial glacial isostatic adjustment of relative sea level (GIA-RSL) in mm/year (used to fuse with GIA-VLM dataset), (c) spatiotemporal satellite altimeter (SA) dataset around North America in mm, and (d) spatiotemporal tide gauge (TG) dataset around North America in mm (used to fuse with SA dataset). Additionally, the spatial TG dataset, as shown in Figure 5.1b, and the spatial SA dataset, as shown in Figure 5.1c, were also used in the experiments.

the parameters from all sources of the datasets for the DF-NSGP, or to use multiple nearest neighbor regions of the surrogate sources.

5.3 Application to Climate Datasets

The experimental setup for the climate-related application of sea-level changes was similar to Section 4.5.1. However, additional data setup for the surrogate sources were also used. For the synthetic dataset of GIA-VLM, the surrogate source of the glacial isostatic adjustment of relative sea level (GIA-SL) was used, as shown in Figure 5.3 a

Table 5.1: Evaluation of various GP models for the spatial and spatiotemporal climate-related datasets at global and regional extents. The spatial datasets used are the glacial-isostatic adjustment (GIA) and tide gauge records (TG), and the results are in the units of mm/year. The spatiotemporal dataset used is the satellite altimeter observations (SA), and the results are in the units of mm. The mean of posteriors from the GP models of SGP (using stationary covariance), NSGP (using non-stationary covariance), HD-NSGP (using high-dimensional manifold non-stationary covariance), and DF-NSGP (using data-fused high-dimensional manifold non-stationary) are compared using the evaluation metrics of MSE and NLPD. The column with the bold values, i.e., DF-NSGP, shows the best evaluation metric score among the various GP models.

Dim.	Datasets	SGP		NSGP		HD-NSGP		DF-NSGP	
		MSE	NLPD	MSE	NLPD	MSE	NLPD	**MSE**	**NLPD**
Spatial ($m = 2$)	GIA (Global)	0.58	3.08	0.56	2.00	0.54	1.94	**0.52**	**1.91**
	GIA (Barents Sea)	1.57	9.23	1.24	7.30	1.03	6.90	**0.92**	**5.70**
	TG (Global)	0.85	2.81	0.75	2.80	0.71	2.78	**0.66**	**2.75**
	TG (N. America)	1.05	3.0	0.54	2.62	0.50	2.58	**0.46**	**2.55**
	TG (S. Africa)	2.10	2.95	2.73	3.18	2.69	3.17	**1.82**	**2.88**
Spatio-temporal ($m = 3$)	SA (N.America)	1.38	4.29	1.18	4.25	1.10	4.15	**1.08**	**4.11**

and b. For the spatial TG dataset, the surrogate source of spatial SA was used, as shown in Figure 5.1b. Finally, for the spatiotemporal dataset of TG, the surrogate source of spatiotemporal SA was used, as shown in Figure 5.3 c and d.

In Table 5.1, we compare various GP models, including SGP, NSGP, HD-NSGP, and DF-NSGP, for the aforementioned spatial and spatiotemporal datasets. In this table, we see that the proposed DF-NSGP model, as described in this chapter, outperforms other GP models including the HD-NSGP model, as described in the previous chapter. One reason for the DF-NSGP model to outperform the HD-NSGP model is that the climate-related measurements around certain regions are spatially sparse. For example, the spatially sparse and irregularly spaced TG sites around South Africa and North

America show a higher order of improvement in Table 5.1 for the DF-NSGP model compared to the other datasets. Hence, using multiple sources of related processes for modeling can particularly be useful when the measurements are insufficient.

5.4 Discussions

It is well known that estimating the parameters of the non-stationary process requires a lot more samples of the measurements than the stationary process. Furthermore, many of the climate-related processes, including sea-level changes, temperature, and precipitation, are known to exhibit non-stationarity and insufficient measurements. To improve the estimates of the non-stationary process from an existing insufficient sample of measurements, in this chapter, we proposed a data-fused high-dimensional manifold non-stationary covariance function (DF-NSGP) and its parameter estimation scheme for the non-stationary GP model. The primary focus of the DF-NSGP model, as presented in this chapter, was to demonstrate the improved estimates of the spatially-varying geometric anisotropy, or scale matrix parameter, for inference of a process. However, the DF-NSGP model can be extended to estimate other parameters of the process, such as the signal and noise variances. In our experiments for climate-related processes, such as sea-level changes, we showed that the DF-NSGP model, when using an additional source of information, outperformed our previously proposed non-stationary model (HD-NSGP), which used a single source of measurement (in Chapter 4). The results especially improved for the regions where there were insufficient, irregular, and spatially sparse records of the climate variable, such as South Africa and North America.

Alternatively, a more traditional approach to improve inference is to use covariates, or features, in the model from related processes, or climate variables. See [107] for examples of a non-stationary GP model with the covariate information. However, adding covariates could potentially increase the number of parameters and the computational complexity of the GP model. On the other hand, by using DF-NSGP we can independently and parallelly estimate the parameters of the multiple sources of datasets. Thereby, we render an unchanged size in the number of parameters and computational complexity of the GP model.

In our methodology for data fusion, we assumed that the SPD matrices from two sources of datasets are in the local vicinity on a manifold. Further improvement in the data-fusion model could relax this assumption and provide a scheme to compute the scale matrix when the SPD matrices from the multiple sources are far apart. One such potential scheme could be to use the Schild's ladder algorithm (see [84]) to compute the distances, and, thereby, its intrinsic statistics to estimate the scale matrix parameters.

In our experimental pilot setup, we considered the climate model output (i.e., GIA-LVM and GIA-RSL), remote sensing dataset (i.e., satellite altimeter measurements), and station data (i.e., tide gauge stations). Given the promising results using the DF-NSGP model from only two sources of datasets, future work could include multiple related sources of datasets to improve our understanding of particular climate phenomena.

Chapter 6

Intermodel Comparison and Emulation Geostatistical Model

Each group has something to say, something about the future of our Earth. I am not interested in a group or its words, but, in the theory of bringing them all together to see the future of our Earth clearly . . . can we?

6.1 Introduction

Various groups worldwide provide us with their Earth system models (ESM) that describe the Earth's physical, biological, and chemical processes. These ESMs typically use solutions to the differential equations of thermo-fluid dynamics to realize space and time dependent values of various physical processes. One of the primary goals of ESMs is to simulate future Earth scenarios. For example, Figure 6.2 shows the future scenario of precipitation changes across North America from three different groups, i.e., Max Planck Institute in Germany, Meteorological Research Institute in Japan, and Community Earth System Model in the USA. Some of these groups show us a drastically different view of the future climate change. On the other hand, due to inherent dependencies in the legacy codes and data of ESMs, there also exists similar patterns from ESMs output. For example, in Figure 6.2 one can visually see the similarities in the patterns of precipitation changes around North America from two of the ensemble members, i.e., NorESM1-M and NorESM1-ME. For the known intermodel dependencies that are based on legacy code and data sharing, see the genealogy map in [21].

An ongoing project of the Coupled Model Inter-comparison Project (CMIP [49]) coordinates efforts between various Earth system modeling groups to create a database of multi-model ensembles of climate simulations. For our experimental setup, we use

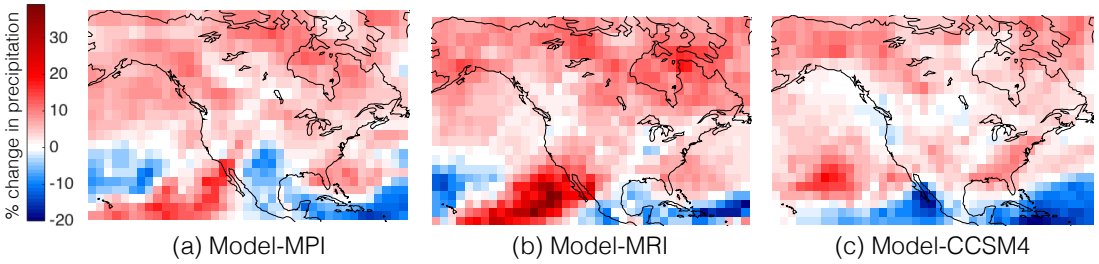


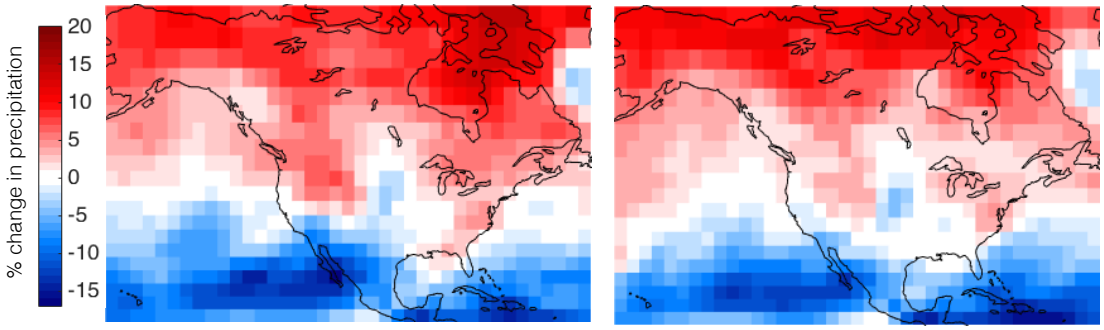
Figure 6.1: Projections (2090) of percent change in precipitation per degree of change in the global mean temperature for North America from the Climate Model Intercomparison Project-5 (CMIP5) ensemble. Shown here are projections from the (a) Max Planck Institute (MPI, Germany), (b) Meteorological Research Institute (MRI, Japan), and (c) Community Earth System Model (CCSM4, USA).

the ESM outputs that were assembled for the fifth phase of this project (CMIP5).

In Section 1.3, we have described several challenges in the CMIP5 ensemble. Furthermore, in this section we also describe our motivation for developing the intermodel comparison and emulation geostatistical models on a Riemannian manifold of the ESM's covariance structure. A framework to compare models is called an intermodel comparison model. Similarly, a framework to statistically simulate Earth's future realizations is called an emulator.

An overview of various methodologies and their limitations for capturing the variability among climate model outputs is given in [24]. For example, some of the climate models share common physical representation and numerical methods, and, thereby, cannot be considered as independent simulations. Additionally, the dependencies in climate models reduce the spread of future climate projections. A Bayesian hierarchical framework has been suggested by [51, 52, 53, 54] to address the intermodel dependency issues. However, the proposed Bayesian framework faces difficulties in robustly modeling the inter-dependencies because the posterior distributions are excessively sensitive to prior assumptions. Additional approaches for a specific climate variable based on the geophysics of specific regions have also been compiled in [20]. However, our goal is to provide a general non-parametric framework that has the potential to be applied to any or all of the Earth system components.

A recent work by [64] shares similar methodological goals as ours in that the authors address issues of model dependencies and sampling in a non-parametric framework.



(a) Model: NorESM1-M (b) Model: NorESM1-ME

Figure 6.2: An example of the sample dependence in the CMIP-5 ensemble. Projections (2090) of percent change in precipitation per degree of change in the global mean temperature for North America from the Norwegian Climate Center's Earth System Model (NorESM). The core model from NorESM is named (a) NorESM1-M. Its variation, which includes biogeochemical cycling, is named (b) NorESM1-ME. Note, the spatial patterns in (a) and (b) are visually similar.

The authors use a standard Euclidean metric on a low-dimensional space by fixing the modes of variance within the available ensemble. Thus, the authors limit the amount of variability information that is present in the climate model outputs. While this method would be a sufficient distance measure for some climate variables, such as climate sensitivity projections, the method lacks all of the variability information that could produce unviable samples for other climate variables, such as precipitation change.

Hence, our goal is to provide a framework where we can compare various ESM outputs in their intrinsic high-dimensional ($m \geq 50$) space of covariance structure, and, thereby, emulate the future Earth scenario.

6.2 High-dimensional Manifold Intermodel-Comparison Model

The intuition behind our proposed intermodel comparison model is that the ensemble of a well-fitted covariance matrix from each climate model output has sufficient information to provide us with a distance measure. Using this distance measure, we can viably compare ESM's outputs. As the covariance matrices of the ensemble members lie on the space of a SPD manifold, we can employ the distance function of the geodesic curve on the manifold, as described in Section 2.2.3. In Figure 6.3, we depict our methodology, and, in Scheme 5, we present our model for the intermodel comparison of climate model

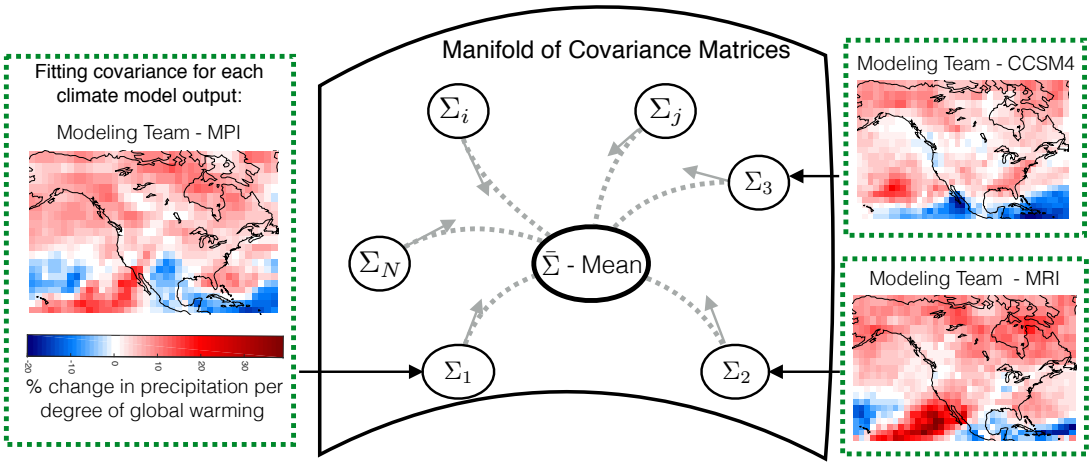


Figure 6.3: A proposed intermodel comparison model using the high-dimensional manifold representation of covariance matrices. Climate model outputs from individual modeling teams of the CMIP5 ensemble, such as CCSM4, MPI, and MRI, are used to parameterize each of the covariance matrices.

outputs.

Scheme 5: Finding the distance function of climate model outputs.

1: Choose an appropriate covariance function for each of the climate variables and its associated climate model output. Then, the covariance matrix is given as follows: $K_i(\theta_i) = \sigma^2_{\boldsymbol{Z}} H(\Sigma_{\boldsymbol{Z}}) + \sigma^2_{\boldsymbol{Y}} \mathbf{I}$, where $i = 1 \dots n$ is the number of ensemble members, and $m = \dim(K)$ is the number of geolocations for which the ESM provides a response, or output. To retain the SPD manifold notation, let $\Sigma_i(\theta_i) = K_i(\theta_i)$.

2: Estimate the parameters of the covariance function for each ensemble member, i.e., $\theta_i = \{\sigma_{\boldsymbol{Y}}, \sigma_{\boldsymbol{Z}}, \Sigma_{\boldsymbol{Z}}\}$, using MLE, i.e., $\theta_i = \text{argmax}_\theta \mathcal{L}(\theta, \mathbf{z_i}(\boldsymbol{x_1}), \dots, \mathbf{z_i}(\boldsymbol{x_m}))$. Here, $\mathbf{z_i}$ is the multivariate response from the ESM output, and m is the number of geolocations.

3: Compute Σ_i by fitting the estimated parameters to the chosen covariance function.

4: Compute the geodesic curve on a manifold of covariance structure, i.e., $D^2(\Sigma_1, \Sigma_2) = \frac{1}{2}\text{tr}(\log^2(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}}))$, where $\Sigma_i \in S^+(m, \mathbb{R})$. For the computational details of the geodesic curve, see 2.2.3.

For simplicity, we used the stationary anisotropic Matérn covariance function in our

experiments. However, the model can be easily extended to the previously proposed covariance functions of HD-NSGP and DF-NSGP, as presented in Chapters 4 and 5, respectively. The geostatistical parameters of the covariance function, i.e., {sill, range, and nugget}, can be estimated using maximum likelihood estimation (MLE). Again, for simplicity, we fix the smoothness parameter of the Matérn function to $\nu = 1$. The previously described MCMC scheme can be plugged in to get the posterior estimates of the geostatistical parameters when using the HD-NSGP or DF-NSGP covariance function.

We compare the validity of our distance function with the genealogy map of [21], and, further, show a contrasting result when compared with the multi-dimensional scaled Euclidean metric framework of [64]. The geodesic distance function also provides us with a metric for the proposed emulator model, as described in the next section.

6.3 High-dimensional Manifold Emulation Model

The core idea of our high-dimensional manifold emulator is to sample the patterns (i.e., covariance structure) of realizations from a distribution, rather than the standard sampling scheme of sampling individual points from the distribution. The standard multivariate sampling scheme (MVN) is described in Section 3.3. To show the applicability of our idea, in this thesis we resort to a multivariate normal distribution for the high-dimensional manifold emulator (HD-MVN). By sampling the entire covariance matrix, $\{\Sigma\}$, from the underlying distribution, which is estimated from existing ensemble members' realizations, we can focus on providing patterns of future climate change. However, the existing ensemble members, as described in Section 6.1, have inherent dependencies. To reduce the bias from the estimates of the existing ensemble members, we employ a weighted sample scheme. Below we describe our proposed emulator

(HD-MVN) for a new sample, $\tilde{\mathbf{z}}$, that represents the future climate projection:

$$\mathbf{Z} \sim \text{MVN}(\boldsymbol{\mu}, \Sigma(\theta)), \quad \tilde{\mathbf{z}} = \boldsymbol{\mu} + \Sigma^{\frac{1}{2}}\boldsymbol{\epsilon}, \tag{6.1a}$$

$$\hat{\boldsymbol{\mu}} = \sum_{j=1}^{N}\sum_{k=1}^{n_j} \frac{1}{Nn_j}\mathbf{z}_{j,k}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{6.1b}$$

$$\Sigma(\theta) \sim \mathcal{N}(\bar{\Sigma}, \Lambda | \Sigma_1(\theta_1), \ldots, \Sigma_{\text{n}}(\theta_{\text{n}})), \quad \hat{\Sigma} = \bar{\Sigma} + \Lambda^{\frac{1}{2}}\boldsymbol{\epsilon}. \tag{6.1c}$$

For simplicity, we assume that the climate model outputs are from the multivariate normal distribution, i.e., $\mathbf{Z} \sim \text{MVN}(\boldsymbol{\mu}, \Sigma(\theta))$. The ensemble mean, $\hat{\boldsymbol{\mu}}$, is estimated as a weighted average of the ESM's output, $\{\mathbf{z_1}(\boldsymbol{x}), \ldots, \mathbf{z_n}(\boldsymbol{x})\}$. Here, N is the number of clusters of ESMs, and n_j is the number of ESM members in each cluster. The ensemble covariance matrix, Σ, is sampled from the normal distribution on the manifold of covariance matrices. Furthermore, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the noise variance vector that is sampled from a standard normal distribution. The parameters $\bar{\Sigma}$ and Λ are the mean and variance, respectively, of the ensemble of covariance matrices, as given in Section 2.3.2.

Similar to Section 6.2, each $\Sigma_i(\theta_i)$ is a covariance matrix of individual ensemble members, and θ_i is estimated using the MLE, i.e., $\theta_i = \text{argmax}_\theta \mathcal{L}(\theta, \mathbf{z_i}(\boldsymbol{x_1}), \ldots, \mathbf{z_i}(\boldsymbol{x_m}))$. Note, there is a difference in the parameter estimation for the MVN and HD-MVN model. In MVN, all ensemble members, and their input space points, are used to estimate the parameters, while in HD-MVN, the parameters are estimated parallelly and separately for each ensemble member. Hence, HD-MVN has a computational advantage over MVN.

Additionally, in HD-MVN the weights of the samples are informed by clustering similar covariance matrices on a manifold; this is in contrast to the MVN model. For example, the mean estimates in the standard MVN model have weights that are either constant or informed by a prior distribution of the input space points. In Figure 6.4, we depict the effects of clustering, where the similar ESM members, as seen in Figure 6.2,

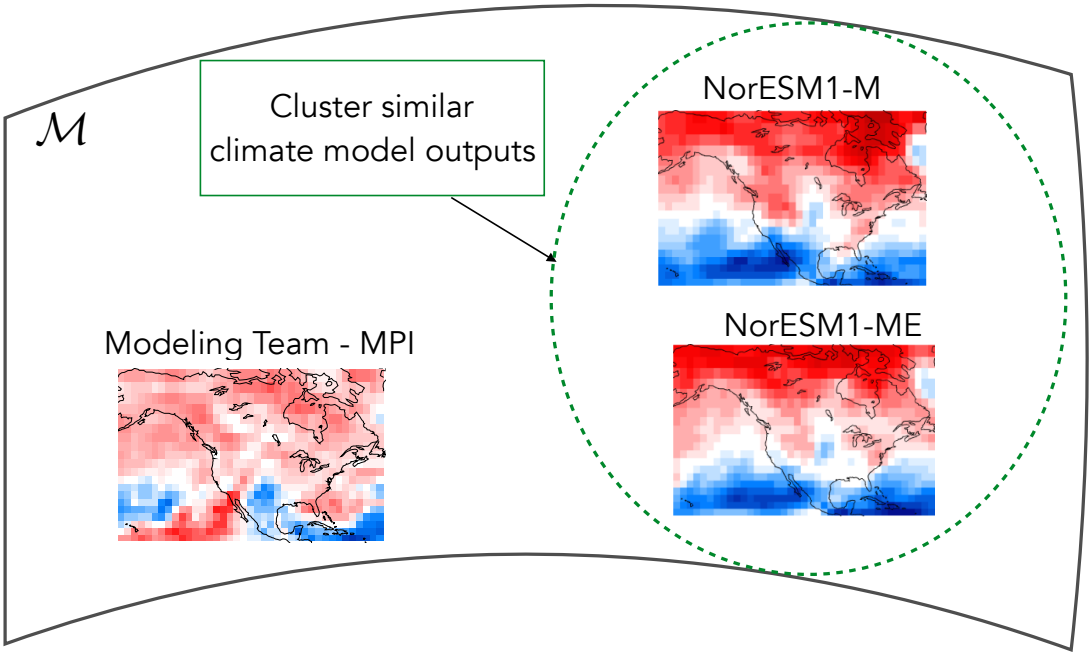


Figure 6.4: An example of the climate model output cluster on a high-dimensional manifold of covariance matrices. Such clusters inform the weights for the estimation of the mean parameter of the sample distribution.

were clustered because they were close on a manifold of covariance structure. To cluster the covariance matrices on a manifold, we used a standard hierarchical clustering method with a criteria as follows: $\max\{D(\Sigma_1, \Sigma_2) : \Sigma_1 \in N_1(\Sigma_i), \Sigma_2 \in N_2(\Sigma_j)\} < \tau$. The threshold, τ, was empirically chosen in our experiments. Here, N_1 and N_2 are two sets of clusters, and the geodesic on the manifold of covariance matrices was given in Section 6.4. Intuitively, the estimates of $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ in the HD-MVN model incorporate extra information about the structure of covariance matrices that the MVN fails to consider. In Appendix B.2.2, Proposal Scheme 9, we outline the details of the HD-MVN sampling procedure.

We validate our clustering method with the genealogy map of [21]. Furthermore, we compare the geostatistical parameters from our new realizations with the MVN model's realizations using the experimental semi-variogram plots. Given the semi-variogram function, one can estimate the parameters (range, sill, and nugget) of the covariance function. Hence, the semi-variogram plot explained in detail in [102] is a good tool in spatial statistics to visualize the differences in covariance matrices. The geostatistical

tool of the semi-variogram plot is briefly described in Section 3. Finally, the patterns of the new realizations are visually compared with the ensemble members to intuitively see the validity of our proposed emulator for climate-related variables.

6.4 Application to Climate Datasets

6.4.1 Experimental Setup

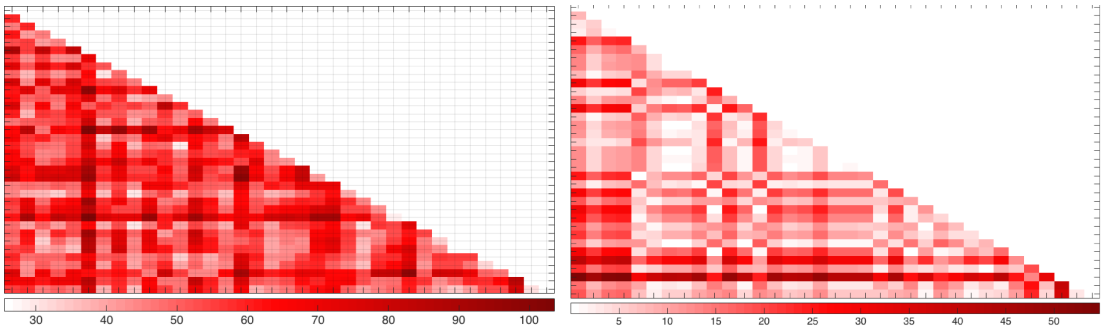
To gain insight into the applicability of our proposed intermodel comparison and statistical emulation method, we used the ensemble of climate model outputs from CMIP5 experiments of future (2090) projections under various representative concentration Pathways (RCP). For the experimental design and different Earth system modeling groups included in CMIP5, see [49]. To test our method against various patterns in climate model outputs, we selected one of the most poorly understood and simulated climate variables: percent change in precipitation per degree of change in the global mean temperature.

In this thesis, we restrict our study to the spatial dataset of the North American region to analyze the regional spatial variability aspect of the climate model outputs. Additionally, we have included single simulation runs from each of the ESMs, rather than multiple simulations, to reduce biases in the ensemble. In Section 1.3, we provide more details of the CMIP5 ensemble size and the RCP scenarios.[1]

6.4.2 Results

In Figure 6.5, we compare our proposed distance function (i.e., geodesic distance) for intermodel comparison, as described in Section 6.2, and the multi-dimensional scaled Euclidean metric, as described in [64]. In this figure, the rows and columns of the distance matrix plot represent the various ensemble members from CMIP5. Furthermore, the darker the red color in each cell the further apart they are on the chosen distance function's space. The validity of the distance function can be implied by visualizing the contrast in the two subfigures, i.e., Figure 6.5a and b. For example, Figure 6.5a

[1]The CMIP5 datasets are available at `https://cmip.llnl.gov/cmip5/data_description.html`



(a) Euclidean Distance (b) Geodesic Distance

Figure 6.5: The distance matrix for the CMIP5 ensemble using the (a) Euclidean distance between climate model outputs and (b) geodesic distance between covariance matrices of climate model outputs.

shows little contrast across the cells, i.e., mostly dark red, implying that most of the ensemble members are far from each other on the Euclidean space. On the other hand, Figure 6.5b shows high contrast across the cells, i.e., some dark reds and some light reds, implying that numerous dependencies were identified on the manifold of covariance structure.

We further validate our proposed distance function (i.e., geodesic distance) for intermodel comparison in Figure 6.6. The climate model outputs from the same Earth system modeling group are highlighted by the blue boxes and are known to have high intermodel dependencies for reasons that include code and data sharing (see [21]). The highlighted blue boxes show lighter shades of red, and, in turn, demonstrate that the chosen geodesic distance function effectively compares and clusters climate model outputs in a non-parametric fashion.

To show the validity of the emulator's new realizations of future projection for various RCP scenarios, we use the experimental semi-variogram plots, as shown in Figure 6.7. In this figure, the red lines represent the semi-variogram fit of the ESM's output, and the blue lines show the emulator's realizations. The spread, depicted with an ellipse, from the HD-MVN model's realizations (Figure 6.7 b, d) is wider than the MVN model's realizations (Figure 6.7 a, c), especially for the RCP scenario of 2.6 and 4.5. For RCP 8.5, the spread from both methods,i.e., HD-MVN (Figure 6.7f) and MVN(Figure 6.7e), are similar, because the underlying ensemble members have lower

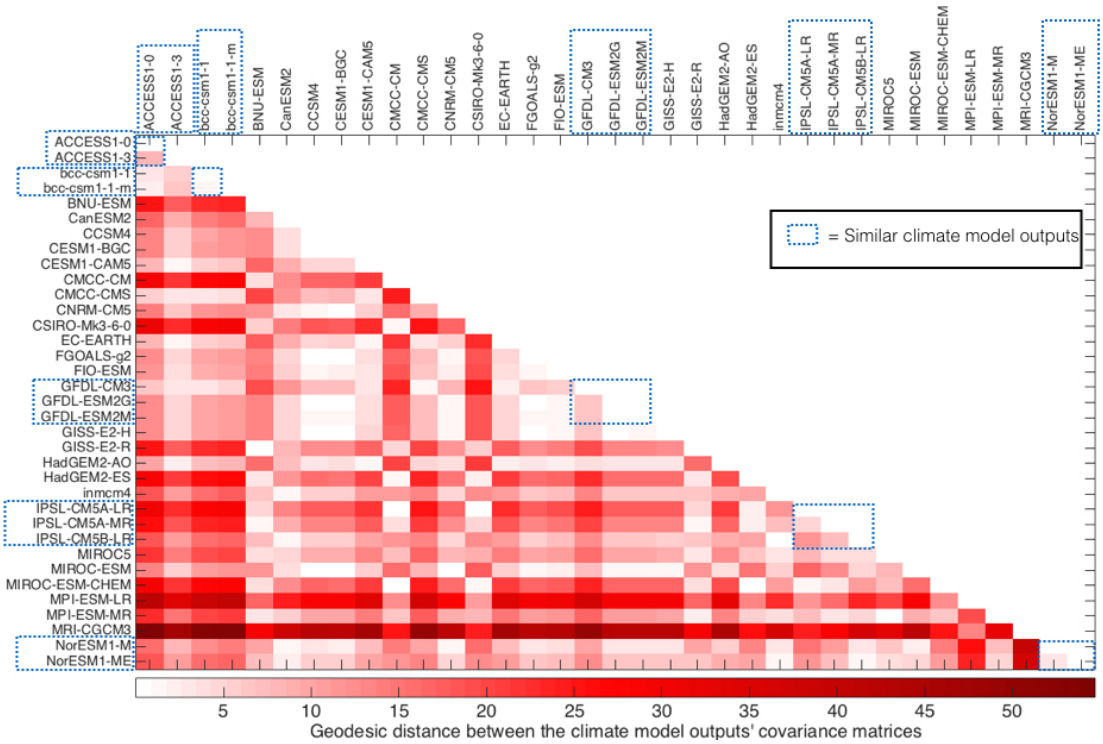


Figure 6.6: A representation of the proposed distance measure between climate model outputs of the CMIP5 ensemble members. The proposed distance measure is a geodesic distance between the fitted covariance matrix of individual climate model outputs. Rows and columns of the above plot represent various climate model outputs from the CMIP5 ensemble. Lighter shades of red represent higher similarity between models, and boxes represent climate models that are validated to have high intermodel dependencies.

intermodel variability when compared to the other RCP scenarios. The wide spread in the HD-MVN samples could be attributed to the sampling of the covariance matrices from a manifold. Furthermore, HD-MVN's realizations encompass the entire spread of the underlying CMIP5 ensemble members. In comparison, the realizations from MVN are small and encompass only a limited number of ESM members from the underlying CMIP5 ensemble members' spread. One reason for the MVN's limited spread could be the equally weighted mean estimate. Hence, implying that the HD-MVN model represents the underlying distribution of the ensemble members better than the MVN model.

In Figure 6.8, we further evaluate the spatial patterns of the emulator's realizations with respect to one of the CMIP5-RCP2.6 ensemble members, i.e., Geophysical Fluid Dynamics Laboratory's climate model output of GFDL-ESM2G. To compare the spatial

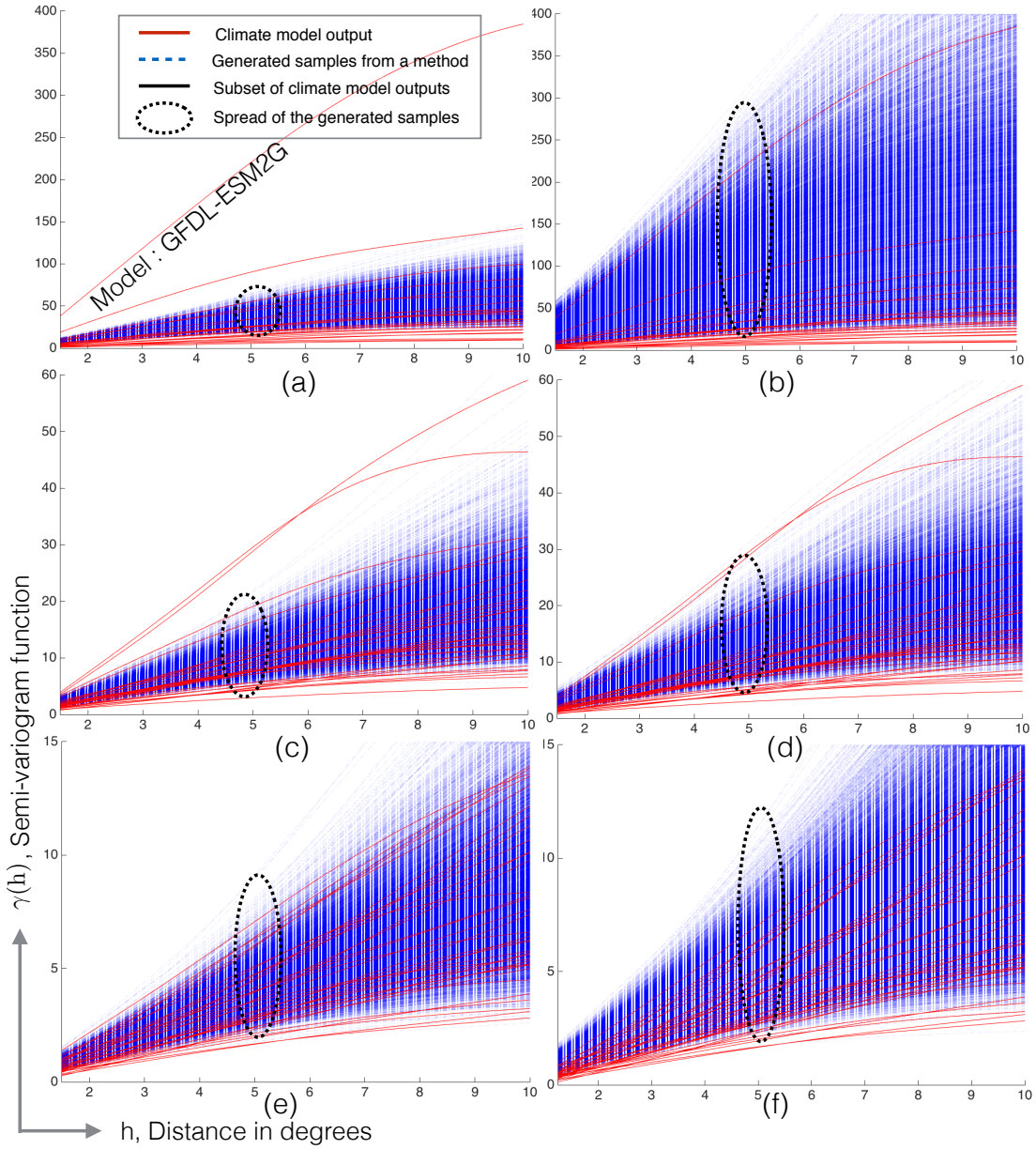


Figure 6.7: Diagnostic plots showing the experimental semi-variogram function for various climate model outputs from CMIP5 ensembles (red lines) and statistically generated samples (blue lines) from the standard multivariate normal sampling method (MVN) for the (a) RCP2.6 ensemble, (c) RCP4.5 ensemble, and (e) RCP8.5 ensemble. Realizations from the proposed high-dimensional manifold sampling method (HD-MVN) are shown for the (b) RCP2.6 ensemble, (d) RCP4.5 ensemble, and (f) RCP8.5 ensemble. The ellipse in each plot focuses on the spread of the generated samples from each sampling method.

patterns from each of the emulators, we chose the closest realization to GFDL-ESM2G

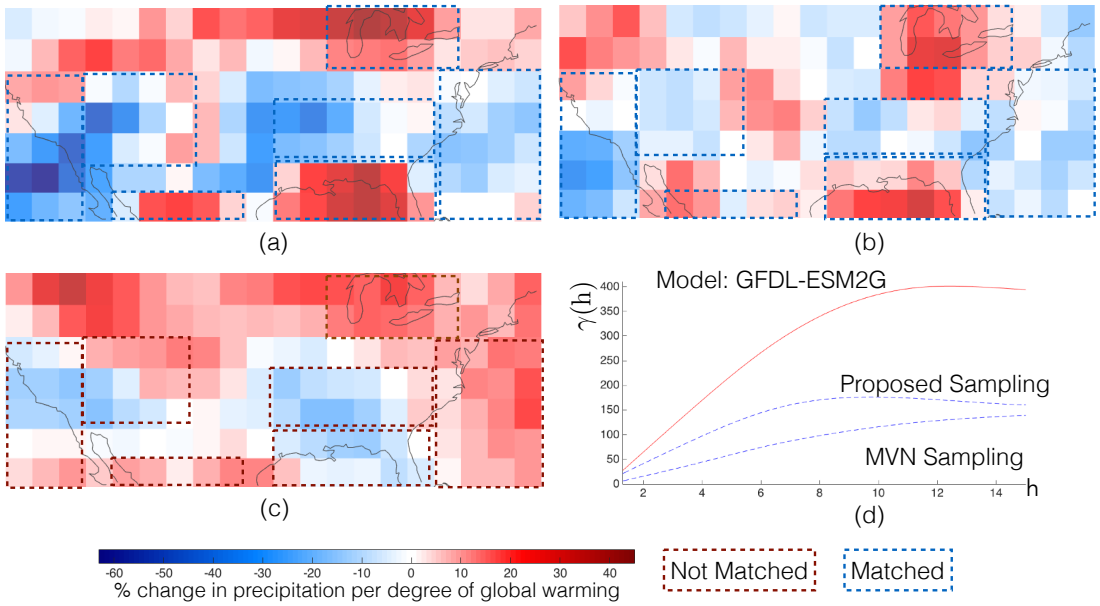


Figure 6.8: Diagnostic plots show the spatial field of climate variables from the Geophysical Fluid Dynamics Laboratory's climate model output of GFDL-ESM2G (a CMIP5-RCP2.6 ensemble member). The spatial field shown here is restricted to the North American region. The figure shows the (a) climate model output from GFDL-ESM2G, (b) closest realization (from Fig. 6.7(b)) using the high-dimensional manifold sampling method (HD-MVN), (c) closest realization (from Fig. 6.7(a)) using the multivariate normal sampling method (MVN), and (d) experimental semi-variogram plots associated with the depicted spatial fields. The coast is represented by black lines, and the boxes represent patterns of similarity between the realizations and the climate model output.

in the semi-variogram plot (Figure 6.8d) from among the thousands of generated realizations. Furthermore, the boxes in HD-MVN (Figure 6.8b) and MVN (Figure 6.8c) represent the matching or non-matching patterns with respect to GFDL-ESM2G (Figure 6.8a). From visual inspection, we see that HD-MVN has more matching spatial patterns when compared to MVN. Also, the semi-variogram fit from HD-MVN is closer to GFDL-ESM2G than MVN. Hence, the aforementioned findings would indicate that HD-MVN's realizations are a better representation of the future projections' spatial patterns than the traditional approaches of MVN, and, thereby, it would be worth pursuing our proposed emulation method to compare and sample climate model outputs.

6.5 Discussion

In this chapter, we proposed an approach to compare and emulate Earth's future projections using the high-dimensional manifold representation of the existing ESM ensemble members. Application of our method also finds its use in the problem of pattern scaling (see [25]). Furthermore, in the experiments, we showed improvements from our proposed approach when compared to the more traditional approaches of comparing and simulation, such as Euclidean distance and MVN sampling scheme, respectively.

Due to the construction of the HD-MVN model, we also computationally benefit from the parallel parameter estimation scheme, which can be useful when considering global samples. To limit the scope of our implementation, we constructed a stationary anisotropic Matérn covariance structure. However, it is straightforward to construct the covariance structure using our previously proposed non-stationary covariance function of HD-NSGP or DF-NSGP.

In our pilot experimental setup, we compared regional variability for precipitation changes from a single global climate model (GCM) run of each ESM ensemble member. However, there are multiple GCM runs for some of the ESMs in CMIP5. For example, the CCSM4 model from NCAR provides several climate model outputs for various RCP scenarios. These GCM runs have varying initial conditions, or perturbations. As we were focused on the decadal scale regional precipitation variability in our experiments, the spatial field from the various GCM runs had little variation. For completeness, it would be useful to consider the multiple GCM runs on the manifold of covariance structure. This experimental setup can provide diagnostics, such as identifying redundancies, by determining the intermodal and intramodal dependencies.

Interesting future work would be to analyze other climate-related variables' future projections from the ESMs' output, such as temperature and climate sensitivity, either separately or on a common manifold of covariance structures. Finally, one of our limitations in the methodological development was the unavailability of ground truth and the small ensemble size. Moving forward, to successfully develop machine learning

approaches to compare and emulate, it would be useful to primarily focus on advancing the assessment procedures. Through this effort to develop a robust emulator and comparison model, we can provide a comprehensive understanding of the uncertainties in future climate projections.

Chapter 7

Concluding Remarks

7.1 Summary

In this thesis, we showed that modeling objects, such as geometric anisotropy and the realizations of our future earth, on a high-dimensional manifold, can potentially provide new insights for climate-related problems. In many fields, such as biomedical imaging, the manifold construction has already shown promising results. However, in climate modeling, this thesis provided novel contributions in the application of high-dimensional manifolds to geostatistical estimation and simulation models. In particular, we introduced high-dimensional manifold geostatistical models for the non-stationary covariance function, data-fused covariance function, intermodel comparison, and a statistical emulator.

In Chapter 4, we provided an extension to a popular class of kernel convolution non-stationary covariance functions (NSGP) (see [7, 8]). The extension was provided by modeling the covariance function's geometric anisotropic parameter on a Riemannian manifold (HD-NSGP). The NSGP model, in [8], used GPs to estimate each of the parameters of an eigendecomposed geometric anisotropy matrix. This modeling approach caused several issues including an over-smoothed realization, over-parameterized model, and rapidly growing (w.r.t. the covariate space dimension) hyperparameters. To address these issues, we modeled the SPD scale matrix in its intrinsic space, which is endowed with a Rao's Riemannian distance. To estimate the parameters of our model we also provided suitable MCMC schemes. Furthermore, we proved that using these intrinsic statistics preserves the smoothness and mean squared differentiable properties of the non-stationary Gaussian process.

By modeling the non-stationary covariance function's parameters in its intrinsic

space, we enabled the spatially-varying scale matrix hyperparameters to grow in proportion to the dimension of the covariate space. One could also employ clustering of geolocations based on geophysics, or priors, to further reduce the number of hyperparameters. Intuitively, such an HD-NSGP model allows for flexibility in the non-stationary model by inducing regional stationarity in the stochastic process. A regionally non-stationary model could be favorable when certain geolocations have few near-by records, or there exists a dominant physical process influencing the specific region. Such situations are common in many climate-related variables, such as sea-level changes.

To show the applicability of the HD-NSGP model, we created a novel simulation function to generate a regionally-varying non-stationary process. Besides, we evaluated various GP models on the spatial and spatiotemporal sea-level change datasets. In our simulation study, we show that the MCMC chain from the HD-NSGP model mixes better and faster than the NSGP model. Also, HD-NSGP was able to recover the regional geometric anisotropy estimates, while NSGP provided an over-smooth estimate. For both the simulation and sea-level datasets, we show that HD-NSGP outperforms the NSGP model for various evaluation criteria. Finally, we also show that the estimation of the geometric anisotropy clusters from the HD-NSGP model conforms with the known geophysical regions around North America.

To further improve the estimates, primarily, when there are few near-by records, we provided a data-fused non-stationary covariance function (DF-NSGP) (see Chapter 5). The DF-NSGP model improves the estimates of the covariance function's parameters, such as geometric anisotropy, by leverages multiple sources of associated measurements. The fusion technique we provided is an extension of our high-dimensional geostatistical modeling approach wherein the geometric anisotropy from multiple datasets are estimated on a manifold of SPD matrices. Similar to the HD-NSGP model, we proved that DF-NSGP preserves the smoothness and mean squared differentiability properties of the non-stationary Gaussian process. Furthermore, for the sea-level change datasets, we showed that DF-NSGP outperforms various GP models that use single-sourced datasets.

Next, in Chapter 6, we showed how the manifold approach can provide valuable

diagnostic tools for high-dimensional climate-related problems. In particular, we provided a distance function to find the dependencies in the various Earth system models' (ESM) future projections, or outputs. After establishing a manifold representation for the ESMs' outputs, we provided a novel statistical emulator model (HD-MVN). The HD-MVN model is capable of generating an unbiased estimate from the ensemble of ESMs' outputs. For validating our distance function and the emulator model, we ran experiments on the future (2090) projection of precipitation changes from the CMIP5 ensemble and for various RCP scenarios. Due to the nature of the problem, we have no ground truth and small ensemble size. Hence, we compared our approach of considering the covariance structure on a manifold with a more traditional previous approach of point-wise modeling in the Euclidean space. For the intermodel comparison, we showed that our distance function detects more of the known dependencies when compared to the Euclidean distance. Similarly, when compared to a multivariate sampling (MVN) model, the HD-MVN model provided a better representation of the future projection of precipitation changes around North America.

In conclusion, this thesis provides original contributions in developing the geostatistical models for estimations and simulations. Furthermore, it provides a novel application of intrinsic statistics on a Riemannian manifold for geostatistical models. Using sea-level and precipitation change datasets, we showed the potential of the high-dimensional manifold geostatistics approach to improve our understanding of physical processes.

7.2 Shortcomings and Potential Improvements

In Sections 4.6, 5.4, and 6.5, we described specific shortcomings and suggested future improvements relating to the proposed high-dimensional geostatistical models and their applications to climate-related datasets. Here, we indentify several shortcomings and suggest potential future directions that are common to all of the models proposed in Chapters 4, 5, and 6.

One of the major challenges with spatial models, including the Gaussian process

(GP), is their inability to handle large datasets. Recently, in the spatial statistics literature, there has been a significant development in modeling for large datasets. Some of the popular stationary spatial models for large datasets are the Lattice Kriging (see [96]), Nearest Neighbor GP (see [97]), and stochastic PDE (see [118]). The main idea to reduce the computational complexity in some of these spatial models is to model the precision matrix, directly, instead of the covariance matrix. The development of non-stationary spatial models for large datasets is a relatively new and ongoing research area within the spatial statistics community. In advancing our non-stationary modeling approach for large global datasets, it would be useful to estimate the regionally-varying precision matrix parameters, for the aforementioned spatial models, on a manifold of high-dimensional geometric structures.

Another aspect of the model development is to provide an estimation procedure that uses the likelihood-based (MLE) schemes rather than resorting to the MCMC scheme. MLE-based schemes are especially useful when considering inference from large datasets and assessing uncertainties in the parameter estimation. A promising likelihood-based estimation procedure for the non-stationary spatial model can be found in [113]. Our non-stationary GP model was inherently based on the covariance structure defined on two separate spaces: one on the input covariate space of the process and the second on the manifold of geometric anisotropy object. Hence, a straightforward closed-form solution of the likelihood function was unavailable. Hence, we resorted to a rather ad-hoc MCMC scheme, which provided satisfactory results for both sampling the high-dimensional objects from a distribution on a manifold and the posteriors of the GP model. The intrinsic statistics of the first-order and second-order geometry on the matrix manifold can be further optimized using the schemes presented in [85]. However, a useful future exercise would be to work out a closed-form solution of the HD-NSGP model to provide MLE-based schemes.

In this thesis, we primarily considered the canonical approach for the intrinsic statistics on the manifold. All the same, there are many facets of the manifold construction, such as the metric, connection, parallel transportation, and vector space, that can be further improved computationally and theoretically. For example, the affine-invariant

metric, or Rao's Riemannian metric, which we used in all of our geostatistical models, involves the expensive computational operation of matrix multiplications. An alternate, computationally inexpensive, and approximate Riemmaian metric to circumvent the matrix multiplications is proposed in [10]. This metric is also called the log-Euclidean metric and is given as follows: $D(\Sigma_1, \Sigma_2) = \| \log(\Sigma_1) - \log(\Sigma_2) \|_F^2$. Furthermore, we assumed a Levi-Civita connection of the tangent spaces a manifold. A potentially useful assumption could be the α-connection (see [76]), or an algorithmically determined arbitrary connection (see [84]), or a connection for discrete surfaces (see [142]). It would be worth exploring various connections on a manifold and their effect on the estimation of spatially-varying parameters for complex spatial processes.

Similarly, analyzing the effects of various vector spaces that represent a compactly smooth manifold could potentially improve our HD-NSGP model. In our framework, we restricted our algebraic computation on the tangent spaces of a canonical Riemannian manifold. In the machine learning community, efficient vector spaces, such as the Grassmannian manifold and Steifel manifold (see [79]), have shown promise in representing objects in their intrinsic high-dimensional space. In our future work, we could explore such vector spaces and tailor them specifically to climate-related variables under consideration.

In climate science, the observational datasets or records are unlike data in other fields where the ground truth can be generated, such as in computer vision. For example, we can create many variations of cat images for predicting objects in an image; however, for predicting sea-level changes, the available tide gauge records are fixed and limited. As a result of the relatively small amount of data available, numerous popular machine learning methods, including Deep Learning, are yet to prove their effectiveness for problems in climate science. Furthermore, due to the small data size in climate science, the simulation study is an essential aspect of improving the models for prediction. To show the validity of our approach for regionally-varying smooth process parameters, we proposed a simulation study in Section 4.4. However, an extensive simulation study that considers many more aspects of the complex climate-related processes, such as teleconnection patterns (see [143]), could improve global spatial models. A useful

future exercise would be to extend the simulation function and the HD-NSGP model to comply with several known geophysics of the climate process. For example, we can algorithmically cluster regional parameters with the neighborhood defined on the manifold and the input space (i.e., geolocations) to identify the teleconnection patterns.

A relatively straightforward extension of our HD-NSGP model could include estimating the regionally-varying and anisotropic noise and signal variance on a manifold. In DF-NSGP, we showed a data-fusion method from two sources of datasets. This model is flexible enough to include multiple sources of datasets. Furthermore, for the intermodel comparison and emulation model, a useful application that could improve our future projections of the climate variables would be to include multiple global climate models (GCM) runs, as well as numerous climate-related variables, on a common manifold construction of climate model outputs. In general, all of the models presented in this thesis is flexible enough to be applied to other climate-related variables, such as precipitation, temperature, and climate sensitivity, for their estimation and simulations.

Thinking of geostatistics as points . . . I suppose can only take us so far. However, thinking about geostatistical objects in high-dimensional spaces can surely provide us with new insights and understanding, as attempted in this thesis. Regardless, can we, or AI, truly comprehend the changes in the natural phenomena?

References

[1] P. J. Diggle and P. J. Ribeiro, *Model-based Geostatistics*. Springer Series in Statistics, 2007.

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

[3] M. Stein, *Interpolation of spatial data: some theory for kriging*. N.Y.: Springer, 1999.

[4] H. Wackernagel, *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media, 2013.

[5] A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, *Handbook of Spatial Statistics*. CRC press, 2010.

[6] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, 2014.

[7] D. Higdon, J. Swall, and J. Kern, "Non-stationary spatial modeling," *Bayesian statistics*, vol. 6, no. 1, pp. 761–768, 1999.

[8] C. Paciorek and M. Schervish, "Nonstationary covariance functions for gaussian process regression," *Advances in neural information processing systems*, vol. 16, pp. 273–280, 2004.

[9] X. Pennec, "Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 1, pp. 127–154, 2006.

[10] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-euclidean metrics for fast and simple calculus on diffusion tensors," *Magnetic resonance in medicine*, vol. 56, no. 2, pp. 411–421, 2006.

[11] S. Amari and H. Nagaoka, *Methods of information geometry*. American Mathematical Soc., 2007, vol. 191.

[12] C. R. Rao, "On the distance between two populations," vol. 21, no. 9, pp. 246–248, 1949.

[13] D. G. Krige, "A statistical approach to some basic mine valuation problems on the witwatersrand," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 52, no. 6, pp. 119–139, 1951.

[14] G. Matheron, "Principles of geostatistics," *Economic geology*, vol. 58, no. 8, pp. 1246–1266, 1963.

[15] R. Schutt and C. O'Neil, *Doing data science: Straight talk from the frontline.* O'Reilly Media, Inc., 2013.

[16] J. G. Dy, J. Emile-Geay, V. Lakshmanan, and Y. Liu, *Proceedings of the Fifth International Workshop on Climate Informatics: CI 2015*, no. 978-0-9973548-0-5, September 2015.

[17] A. Banerjee, W. Ding, J. Dy, V. Lyubchich, A. Rhines, I. Ebert-Uphoff, C. Monteleoni, and D. Nychka, *Proceedings of the Sixth International Workshop on Climate Informatics: CI 2016*, no. 978-0-9973548-1, September 2016.

[18] M. L. Stein, "Nonstationary spatial covariance functions," *Unpublished technical report*, 2005.

[19] P. D. Sampson and P. Guttorp, "Nonparametric estimation of nonstationary spatial covariance structure," *Journal of the American Statistical Association*, vol. 87, no. 417, pp. 108–119, 1992.

[20] IPCC, "The physical science basis: Working group 1 contribution to the fifth assessment report of the intergovernmental panel on climate change," *New York: Cambridge University Press*, vol. 1, pp. 535–1, 2013.

[21] R. Knutti, D. Masson, and A. Gettelman, "Climate model genealogy: Generation CMIP5 and how we got there," *Geophysical Research Letters*, vol. 40, no. 6, pp. 1194–1199, 2013.

[22] R. J. Muirhead, *Aspects of multivariate statistical theory.* John Wiley & Sons, 2009, vol. 197.

[23] W. G. Cochran, *Sampling Techniques: 3rd Ed.* Wiley, 1977.

[24] C. Tebaldi and R. Knutti, "The use of the multi-model ensemble in probabilistic climate projections," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1857, pp. 2053–2075, 2007.

[25] C. Tebaldi and J. M. Arblaster, "Pattern scaling: Its strengths and limitations, and an update on the latest model simulations," *Climatic change*, vol. 122, no. 3, pp. 459–471, 2014.

[26] J. A. Church, P. U. Clark, A. Cazenave, J. M. Gregory, S. Jevrejeva, A. Levermann, M. Merrifield, G. Milne, R. Nerem, P. Nunn *et al.*, "Sea level change," *Climate change*, pp. 1137–1216, 2013.

[27] C. S. Watson, N. J. White, J. A. Church, M. A. King, R. J. Burgette, and B. Legresy, "Unabated global mean sea-level rise over the satellite altimeter era," *Nature Climate Change*, vol. 5, no. 6, p. 565, 2015.

[28] C. C. Hay, E. Morrow, R. E. Kopp, and J. X. Mitrovica, "Estimating the sources of global sea level rise with data assimilation techniques," *Proceedings of the National Academy of Sciences*, 2012.

[29] R. E. Kopp, R. M. DeConto, D. A. Bader, C. C. Hay, R. M. Horton, S. Kulp, M. Oppenheimer, D. Pollard, and B. H. Strauss, "Evolving understanding of antarctic ice-sheet physics and ambiguity in probabilistic sea-level projections," *Earth's Future*, 2017.

[30] J. A. Church, N. J. White, R. Coleman, K. Lambeck, and J. X. Mitrovica, "Estimates of the regional distribution of sea level rise over the 1950-2000 period," *Journal of Climate*, vol. 17, no. 13, pp. 2609–2625, 2004.

[31] R. E. Kopp, C. C. Hay, C. M. Little, and J. X. Mitrovica, "Geographic variability of sea-level change," *Current Climate Change Reports*, vol. 1, no. 3, pp. 192–204, 2015.

[32] J. Church, J. M. Gregory, P. Huybrechts, M. Kuhn, K. Lambeck, M. Nhuan, D. Qin, and P. Woodworth, "Changes in sea level," in *Climate Change 2001: The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel*, 2001, pp. 639–694.

[33] A. C. Kemp, B. P. Horton, J. P. Donnelly, M. E. Mann, M. Vermeer, and S. Rahmstorf, "Climate related sea-level variations over the past two millennia," *Proceedings of the National Academy of Sciences*, vol. 108, no. 27, pp. 11 017–11 022, 2011.

[34] R. E. Kopp, F. J. Simons, J. X. Mitrovica, A. C. Maloof, and M. Oppenheimer, "Probabilistic assessment of sea level during the last interglacial stage," *Nature*, vol. 462, no. 7275, pp. 863–867, 2009.

[35] D. Wuebbles, D. Fahey, K. Hibbard, D. Dokken, B. Stewart, and T. Maycock, "Climate science special report: Fourth national climate assessment," *US Global Change Research Program: Washington, DC, USA*, 2017.

[36] S. J. Holgate, A. Matthews, P. L. Woodworth, L. J. Rickards, M. E. Tamisiea, E. Bradshaw, P. R. Foden, K. M. Gordon, S. Jevrejeva, and J. Pugh, "New data systems and products at the permanent service for mean sea level," *Journal of coastal research*, vol. 29, no. 3, pp. 493–504, 2012.

[37] C. M. Domingues, J. A. Church, N. J. White, P. J. Gleckler, S. E. Wijffels, P. M. Barker, and J. R. Dunn, "Improved estimates of upper-ocean warming and multi-decadal sea-level rise," *Nature*, vol. 453, no. 7198, pp. 1090–1093, 2008.

[38] R. S. Nerem, D. P. Chambers, C. Choe, and G. T. Mitchum, "Estimating mean sea level change from the topex and jason altimeter missions," *Marine Geodesy*, vol. 33, no. S1, pp. 435–446, 2010.

[39] D. P. Chambers, C. A. Mehlhaff, T. J. Urban, D. Fujii, and R. S. Nerem, "Low-frequency variations in global mean sea level: 1950–2000," *Journal of Geophysical Research: Oceans*, vol. 107, no. C4, 2002.

[40] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning.* Springer series in statistics New York, 2001, vol. 1.

[41] J. A. Lee and M. Verleysen, *Nonlinear dimensionality reduction.* Springer Science & Business Media, 2007.

[42] K. E. Trenberth, "Changes in precipitation with climate change," *Climate Research*, vol. 47, no. 1/2, pp. 123–138, 2011.

[43] S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, H. L. Miller *et al.*, "Contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change, 2007," 2007.

[44] G. J. Huffman, R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, and U. Schneider, "The global precipitation climatology project (gpcp) combined precipitation dataset," *Bulletin of the American Meteorological Society*, vol. 78, no. 1, pp. 5–20, 1997.

[45] R. F. Adler, G. J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin *et al.*, "The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present)," *Journal of hydrometeorology*, vol. 4, no. 6, pp. 1147–1167, 2003.

[46] P. R. Gent, G. Danabasoglu, L. J. Donner, M. M. Holland, E. C. Hunke, S. R. Jayne, D. M. Lawrence, R. B. Neale, P. J. Rasch, M. Vertenstein *et al.*, "The community climate system model version 4," *Journal of Climate*, vol. 24, no. 19, pp. 4973–4991, 2011.

[47] G. A. Meehl, G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer, "The coupled model intercomparison project (CMIP)," *Bulletin of the American Meteorological Society*, vol. 81, no. 2, pp. 313–318, 2000.

[48] G. A. Meehl, C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. Mitchell, R. J. Stouffer, and K. E. Taylor, "The wcrp CMIP3 multimodel dataset: A new era in climate change research," *Bulletin of the American meteorological society*, vol. 88, no. 9, pp. 1383–1394, 2007.

[49] K. E. Taylor, R. J. Stouffer, and G. A. Meehl, "An overview of CMIP5 and the experiment design," *Bulletin of the American Meteorological Society*, vol. 93, no. 4, p. 485, 2012.

[50] R. Knutti and J. Sedláček, "Robustness and uncertainties in the new CMIP5 climate model projections," *Nature Climate Change*, vol. 3, no. 4, p. 369, 2013.

[51] C. Tebaldi, R. L. Smith, D. Nychka, and L. O. Mearns, "Quantifying uncertainty in projections of regional climate change: A bayesian approach to the analysis of multimodel ensembles," *Journal of Climate*, vol. 18, no. 10, pp. 1524–1540, 2005.

[52] C. Tebaldi and B. Sansó, "Joint projections of temperature and precipitation change from multiple climate models: a hierarchical bayesian approach," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 172, no. 1, pp. 83–106, 2009.

[53] N. A. Leith and R. E. Chandler, "A framework for interpreting climate model outputs," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 59, no. 2, pp. 279–296, 2010.

[54] R. Furrer, S. R. Sain, D. Nychka, and G. A. Meehl, "Multivariate bayesian analysis of atmosphere–ocean general circulation models," *Environmental and ecological statistics*, vol. 14, no. 3, pp. 249–266, 2007.

[55] D. L. Donoho *et al.*, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*, vol. 1, p. 32, 2000.

[56] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[57] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *International conference on database theory*. Springer, 2001, pp. 420–434.

[58] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*. Springer Science & Business Media, 2007.

[59] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[60] D. P. Van Vuuren, J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. C. Hurtt, T. Kram, V. Krey, J.-F. Lamarque *et al.*, "The representative concentration pathways: an overview," *Climatic change*, vol. 109, no. 1-2, p. 5, 2011.

[61] S. R. Sain, R. Furrer, and N. Cressie, "A spatial analysis of multivariate output from regional climate models," *The Annals of Applied Statistics*, pp. 150–175, 2011.

[62] S. Castruccio, D. J. McInerney, M. L. Stein, F. Liu Crouch, R. L. Jacob, and E. J. Moyer, "Statistical emulation of climate model projections based on precomputed gcm runs," *Journal of Climate*, vol. 27, no. 5, pp. 1829–1844, 2014.

[63] M. Jun, R. Knutti, and D. W. Nychka, "Spatial analysis to quantify numerical model bias and dependence: How many climate models are there?" *Journal of the American Statistical Association*, vol. 103, no. 483, pp. 934–947, 2008.

[64] B. M. Sanderson, R. Knutti, and P. Caldwell, "Addressing interdependency in a multimodel ensemble by interpolation of model properties," *Journal of Climate*, vol. 28, no. 13, pp. 5150–5170, 2015.

[65] J. M. Oller and J. M. Corcuera, "Intrinsic analysis of statistical estimation," *The Annals of Statistics*, vol. 23, no. 5, pp. 1562–1581, 1995.

[66] C. Bingham, "An antipodally symmetric distribution on the sphere," *The Annals of Statistics*, pp. 1201–1225, 1974.

[67] P. E. Jupp and K. V. Mardia, "A unified view of the theory of directional statistics," *International Statistical Review/Revue Internationale de Statistique*, pp. 261–294, 1989.

[68] M. Calvo and J. Oller, "An explicit solution of information geodesic equations for the multivariate normal model," *Statistics and Decisions*, vol. 9, pp. 119–138, 1991.

[69] X. Pennec, "Probabilities and statistics on riemannian manifolds: A geometric approach," Ph.D. dissertation, INRIA, 2004.

[70] N. N. Cencov, *Statistical decision rules and optimal inference*. American Mathematical Soc., 2000, no. 53.

[71] B. Efron, "Defining the curvature of a statistical problem (with applications to second order efficiency)," *The Annals of Statistics*, pp. 1189–1242, 1975.

[72] A. Dawid, "Discussions to efron's paper," *Ann. Statist.*, vol. 3, pp. 1231–1234, 1975.

[73] S. Amari, *Differential geometry in statistical inference*. Inst of Mathematical Statistic, 1987, vol. 10.

[74] S. Amari, *Differential-geometrical methods in statistics*. Springer Science & Business Media, 2012, vol. 28.

[75] H. Nagaoka and S. Amari, "Differential geometry of smooth families of probability distributions," *Univ. Tokyo, Tokyo, Japan, METR*, pp. 82–7, 1982.

[76] S. Amari, "Information geometry on hierarchy of probability distributions," *IEEE transactions on information theory*, vol. 47, no. 5, pp. 1701–1711, 2001.

[77] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras, "Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor mri processing," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 3, pp. 423–444, 2006.

[78] T. Fletcher, "Geodesic regression on riemannian manifolds," in *Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods for Modelling Biological Shape Variability*, 2011, pp. 75–86.

[79] P. Turaga, A. Veeraraghavan, and R. Chellappa, "Statistical analysis on stiefel and grassmann manifolds with applications in computer vision," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[80] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, "Kernel methods on the riemannian manifold of symmetric positive definite matrices," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 73–80.

[81] Z. Wang and B. C. Vemuri, "An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1. IEEE, 2004, pp. I–I.

[82] P. T. Fletcher and S. Joshi, "Riemannian geometry for the statistical analysis of diffusion tensor data," *Signal Processing*, vol. 87, no. 2, pp. 250–262, 2007.

[83] F. Nielsen and F. Barbaresco, Eds., *Geometric Science of Information - Second International Conference, GSI 2015, Palaiseau, France, October 28-30, 2015, Proceedings*, ser. Lecture Notes in Computer Science, vol. 9389. Springer, 2015.

[84] A. Kheyfets, W. A. Miller, and G. A. Newton, "Schild's ladder parallel transport procedure for an arbitrary connection," *International Journal of Theoretical Physics*, vol. 39, no. 12, pp. 2891–2898, 2000.

[85] P. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[86] G. Lebanon *et al.*, *Riemannian geometry and statistical machine learning*. LAP LAMBERT Academic Publishing, 2015.

[87] J. M. Lee, "Smooth manifolds," in *Introduction to Smooth Manifolds*. Springer, 2003, pp. 1–29.

[88] M. Fréchet, "Les éléments aléatoires de nature quelconque dans un espace distancié," 1948.

[89] P. Manfredo, *Differential geometry of curves and surfaces*. Prentice Hall, 1976.

[90] S. Jensen, "An algebraic theory for normal statistical models with unknown covariance," *Recent Developments*, 1977.

[91] L. T. Skovgaard, "A riemannian geometry of the multivariate normal model," *Scandinavian Journal of Statistics*, pp. 211–223, 1984.

[92] S. Said, L. Bombrun, Y. Berthoumieu, and J. H. Manton, "Riemannian gaussian distributions on the space of symmetric positive definite matrices," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2153–2170, 2017.

[93] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on pure and applied mathematics*, vol. 30, no. 5, pp. 509–541, 1977.

[94] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 3, pp. 735–747, 2005.

[95] M. L. Stein, "Limitations on low rank approximations for covariance matrices of spatial data," *Spatial Statistics*, vol. 8, no. C, pp. 1–19, 2014.

[96] D. Nychka, S. Bandyopadhyay, D. Hammerling, F. Lindgren, and S. Sain, "A multi-resolution gaussian process model for the analysis of large spatial data sets." *Journal of Computational and Graphical Statistics*, no. 1061-8600, 2014.

[97] A. Datta, S. Banerjee, A. O. Finley, and A. E. Gelfand, "Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets," *Journal of the American Statistical Association*, vol. 111, no. 514, pp. 800–812, 2016.

[98] C. E. Rasmussen and C. Williams, "Gaussian processes for machine learning. 2006," *The MIT press, Cambridge, MA, USA*, vol. 38, pp. 715–719, 2006.

[99] G. Matheron, "The intrinsic random functions and their applications," *Advances in applied probability*, vol. 5, no. 3, pp. 439–468, 1973.

[100] M. L. Stein *et al.*, "On a class of space–time intrinsic random functions," *Bernoulli*, vol. 19, no. 2, pp. 387–408, 2013.

[101] N. Lawrence, M. Seeger, and R. Herbrich, "Fast sparse gaussian process methods: The informative vector machine," in *Proceedings of the 16th Annual Conference on Neural Information Processing Systems*, no. EPFL-CONF-161319, 2003, pp. 609–616.

[102] N. Cressie, "Statistics for spatial data: Wiley series in probability and statistics," *Wiley-Interscience, New York*, vol. 15, pp. 105–209, 1993.

[103] V. Vapnik, *Statistical learning theory*. Wiley, New York, 1998, vol. 3.

[104] G. Wahba, *Spline models for observational data*. SIAM, 1990, vol. 59.

[105] S. Banerjee, "On geodetic distance computations in spatial modeling," *Biometrics*, vol. 61, no. 2, pp. 617–625, 2005.

[106] R. J. Adler, *The geometry of random fields*. SIAM, 1981, vol. 62.

[107] M. D. Risser and C. A. Calder, "Regression-based covariance functions for non-stationary spatial modeling," *Environmetrics*, vol. 26, no. 4, pp. 284–297, 2015.

[108] J. H. V. Neto, A. M. Schmidt, and P. Guttorp, "Accounting for spatially varying directional effects in spatial covariance structures," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 63, no. 1, pp. 103–122, 2014.

[109] R. E. Kopp, "Does the mid-atlantic united states sea level acceleration hot spot reflect ocean dynamic variability?" *Geophysical Research Letters*, vol. 40, no. 15, pp. 3981–3985, 2013.

[110] D. Damian, P. D. Sampson, and P. Guttorp, "Bayesian estimation of semi-parametric non-stationary spatial covariance structures," *Environmetrics: The official journal of the International Environmetrics Society*, vol. 12, no. 2, pp. 161–178, 2001.

[111] A. M. Schmidt and A. O'Hagan, "Bayesian inference for non-stationary spatial covariance structure via spatial deformations," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 65, no. 3, pp. 743–758, 2003.

[112] Y. Xiong, W. Chen, D. Apley, and X. Ding, "A non-stationary covariance-based kriging method for metamodelling in engineering design," *International Journal for Numerical Methods in Engineering*, vol. 71, no. 6, pp. 733–756, 2007.

[113] E. B. Anderes, M. L. Stein *et al.*, "Estimating deformations of isotropic gaussian random fields on the plane," *The Annals of Statistics*, vol. 36, no. 2, pp. 719–741, 2008.

[114] D. Nychka and N. Saltzman, "Design of air-quality monitoring networks," in *Case studies in environmental statistics*. Springer, 1998, pp. 51–76.

[115] D. Nychka, C. Wikle, and J. A. Royle, "Multiresolution models for nonstationary spatial covariance functions," *Statistical Modelling*, vol. 2, no. 4, pp. 315–331, 2002.

[116] D. M. Holland, N. Saltzman, L. H. Cox, and D. Nychka, "Spatial prediction of sulfur dioxide in the eastern united states," in *Geostatistics for Environmental Applications*. Springer, 1999, pp. 65–76.

[117] T. Matsuo, D. W. Nychka, and D. Paul, "Nonstationary covariance modeling for incomplete data: Monte carlo em approach," *Computational Statistics & Data Analysis*, vol. 55, no. 6, pp. 2059–2073, 2011.

[118] F. Lindgren, H. Rue, and J. 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)*, vol. 73, no. 4, pp. 423–498, 2011.

[119] D. Higdon, "A process-convolution approach to modelling temperatures in the north atlantic ocean," *Environmental and Ecological Statistics*, vol. 5, no. 2, pp. 173–190, 1998.

[120] M. Fuentes and R. L. Smith, "A new class of nonstationary spatial models," Technical report, North Carolina State University, Raleigh, NC, Tech. Rep., 2001.

[121] M. Fuentes, "A high frequency kriging approach for non-stationary environmental processes," *Environmetrics: The official journal of the International Environmetrics Society*, vol. 12, no. 5, pp. 469–483, 2001.

[122] M. Fuentes, "Spectral methods for nonstationary spatial processes," *Biometrika*, vol. 89, no. 1, pp. 197–210, 2002.

[123] W. Kleiber and D. Nychka, "Nonstationary modeling for multivariate spatial processes," *Journal of Multivariate Analysis*, vol. 112, pp. 76–91, 2012.

[124] M. Katzfuss, "Bayesian nonstationary spatial modeling for very large datasets," *Environmetrics*, vol. 24, no. 3, pp. 189–200, 2013.

[125] C. A. Calder, "A dynamic process convolution approach to modeling ambient particulate matter concentrations," *Environmetrics: The official journal of the International Environmetrics Society*, vol. 19, no. 1, pp. 39–48, 2008.

[126] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.

[127] R. M. Neal *et al.*, "MCMC using hamiltonian dynamics," *Handbook of Markov Chain Monte Carlo*, vol. 2, no. 11, p. 2, 2011.

[128] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid monte carlo," *Physics letters B*, vol. 195, no. 2, pp. 216–222, 1987.

[129] G. O. Roberts and O. Stramer, "Langevin diffusions and metropolis-hastings algorithms," *Methodology and computing in applied probability*, vol. 4, no. 4, pp. 337–357, 2002.

[130] I. Murray and R. P. Adams, "Slice sampling covariance hyperparameters of latent gaussian models," in *Advances in neural information processing systems*, 2010, pp. 1732–1740.

[131] O. F. Christensen, G. O. Roberts, and M. Sköld, "Robust markov chain monte carlo methods for spatial generalized linear mixed models," *Journal of Computational and Graphical Statistics*, vol. 15, no. 1, pp. 1–17, 2006.

[132] J. J. Warnes and B. D. Ripley, "Problems with likelihood estimation of covariance functions of spatial gaussian processes," *Biometrika*, vol. 74, no. 3, pp. 640–642, 1987.

[133] H. Zhang, "Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 250–261, 2004.

[134] O. Pasternak, N. Sochen, and P. J. Basser, "Metric selection and diffusion tensor swelling," in *New Developments in the Visualization and Processing of Tensor Fields*. Springer, 2012, pp. 323–336.

[135] I. J. Good, "Rational decisions," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 107–114, 1952.

[136] P. D. Hoff and X. Niu, "A covariance regression model," *Statistica Sinica*, pp. 729–753, 2012.

[137] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

[138] J. Hwang, S. Lay, M. Maechler, R. D. Martin, and J. Schimert, "Regression modeling in back-propagation and projection pursuit learning," *Neural Networks, IEEE Transactions on*, vol. 5, no. 3, pp. 342–353, 1994.

[139] F. Richards, "A flexible growth function for empirical use," *Journal of experimental Botany*, vol. 10, no. 2, pp. 290–301, 1959.

[140] W. Peltier, "Global glacial isostasy and the surface of the ice-age earth: the ice-5g (vm2) model and grace," *Annu. rev. earth planet. sci.*, vol. 32, pp. 111–149, 2004.

[141] F. Nielsen and F. Barbaresco, *Geometric science of information*. Springer International Publishing, 2015.

[142] R. N. Zantout and Y. F. Zheng, "Determining geodesics of a discrete surface," in *Multisensor Fusion and Integration for Intelligent Systems, 1994. IEEE International Conference on MFI'94*. IEEE, 1994, pp. 551–558.

[143] J. M. Wallace and D. S. Gutzler, "Teleconnections in the geopotential height field during the northern hemisphere winter," *Monthly Weather Review*, vol. 109, no. 4, pp. 784–812, 1981.

[144] B. Schutz, *A first course in general relativity*. Cambridge university press, 2009.

[145] V. I. Arnold, *Mathematical methods of classical mechanics*. Springer Science & Business Media, 2013, vol. 60.

[146] Y. Sato, K. Sugawa, and M. Kawaguchi, "The geometrical structure of the parameter space of the two-dimensional normal distribution," *Reports on Mathematical Physics*, vol. 16, no. 1, pp. 111–119, 1979.

[147] C. Dodson and H. Wang, "Iterative approximation of statistical distributions and relation to information geometry," *Statistical inference for stochastic processes*, vol. 4, no. 3, pp. 307–318, 2001.

[148] T. Imai, A. Takaesu, and M. Wakayama, "Remarks on geodesics for multivariate normal models," *2011B-6*, 2011.

[149] W. Förstner and B. Moonen, "A metric for covariance matrices," in *Geodesy-The Challenge of the 3rd Millennium*. Springer, 2003, pp. 299–309.

[150] T. Vincenty, "Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations," *Survey review*, vol. 23, no. 176, pp. 88–93, 1975.

[151] I. Schoenberg, "Positive definite functions on spheres," *Duke Math. J*, vol. 1, p. 172, 1988.

[152] T. Gneiting, "Strictly and non-strictly positive definite functions on spheres," *Bernoulli*, vol. 19, no. 4, pp. 1327–1349, 2013.

Appendix A

Model Formulations

A.1 Ricci Curvature Tensor

In Riemannian geometry, the curvature tensor, $\mathcal{R}$, intuitively measures the extent to which the neighborhood on a manifold deviates locally from the Euclidean space. The curvature tensor is formulated using the parallel transport operations, which, in turn, is defined with respect to the connection on a manifold (see [144, 145]). In our models, we have chosen the canonical connection, i.e., Levi-Civita connection.

To set our notations, let the symmetric positive definite (SPD) space, $S^+(m, \mathbb{R})$, of the covariance matrices, $\Sigma = \{\sigma_{ij}\}_{i \leq j; i,j=1 \ldots m} \in \mathbb{R}^{\frac{1}{2}(m(m+1))}$, be parameterized as follows: $\theta = \{\theta^i\}_{i=1,\ldots,m'}$. Then, the four-indexed Riemannian curvature tensor corresponding to the Levi-Civita connection and Fisher information metric is given as follows:

$$\begin{aligned} \mathcal{R}_{ijkl} &= \mathcal{R}(X, Y, Z, V) \\ &= \frac{1}{4} \operatorname{tr}\big(Y\Sigma^{-1}X\Sigma^{-1}Z\Sigma^{-1}V\Sigma^{-1}\big) - \frac{1}{4} \operatorname{tr}\big(X\Sigma^{-1}Y\Sigma^{-1}Z\Sigma^{-1}V\Sigma^{-1}\big), \end{aligned} \tag{A.1}$$

where X, Y, Z, V is the canonical basis vector field in the direction of Σ. This curvature tensor has 1,296 components, and, for $m = 3$, there are 105 independent components (see [77]).

Another way to describe the curvature tensor is through the sectional curvature tensor, K. Intuitively, the sectional curvature is the curvature of a two-dimensional section of the manifold. An even simpler formulation is given by the Ricci curvature tensor. The Ricci curvature tensor, Ricci is a symmetric $m' \times m'$ covariant tensor and defined as a contraction of the Riemannian curvature tensor, $\mathcal{R}$, with respect to the

Riemannian metric, g. Intuitively, Ricci measures the growth rate of a metric and, popularly, has played an important role in the Einstein field equations. Furthermore, the Ricci scalar curvature, R, is defined as the trace of the Ricci curvature tensor. For $m = 2$, the Ricci scalar curvature determines the sectional curvature tensor, and, for $m = 3$, the Ricci curvature tensor completely determines the sectional curvature tensor. Further details of the Ricci curvature tensor, Riemmanian curvature tensor, sectional curvature, and Ricci scalar for the multivariate normal distribution (MVN) can be found in [71, 91].

The components of the Ricci curvature tensor are expressed as follows:

$$\text{Ricci}_{ij} = \text{R}_{ijkl} g^{kl}, \tag{A.2}$$

where g^{kl} is the inverse of the metic. The Ricci curvature tensor is central to the derivations of the intrinsic statistics on a Riemannian manifold. Furthermore, the probability density function of the generalized normal distribution (see Equation 2.5) includes the Ricci curvature tensor in computing its normalization factor. However, for certain practical purposes, and in high-dimensions ($m > 3$), one can assume a locally flat or Euclidean manifold (see Equation 2.6), or a scalar variance (see Equation 2.7), to circumvent the computational burden of the Ricci curvature tensor.

The components of Ricci curvature tensor can be computed using any symbolic algebraic softwares (such as Mathematica). For a bi-variate, $m = 2$, normal distribution with the parameter space of $\theta = \{\theta^i\}_{i=1,\ldots 5} = \{\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22}\}$, the components of the metric tensor is given as follows (see [146]):

$$[g_{ij}] = \begin{bmatrix} \frac{\theta^5}{\Delta} & \frac{-\theta^4}{\Delta} & 0 & 0 & 0 \\ \frac{-\theta^4}{\Delta} & \frac{\theta^3}{\Delta} & 0 & 0 & 0 \\ 0 & 0 & \frac{(\theta^5)^2}{2\Delta^2} & \frac{-\theta^4\theta^5}{\Delta^2} & \frac{(\theta^4)^2}{2\Delta^2} \\ 0 & 0 & \frac{-\theta^4\theta^5}{\Delta^2} & \frac{\theta^3\theta^5+(\theta^4)^2}{\Delta^2} & \frac{-\theta^3\theta^4}{\Delta^2} \\ 0 & 0 & \frac{(\theta^4)^2}{2\Delta^2} & \frac{-\theta^3\theta^4}{\Delta^2} & \frac{(\theta^3)^2}{2\Delta^2} \end{bmatrix}, \tag{A.3}$$

where the determinant $\Delta = |\Sigma| = \theta^3\theta^5 - (\theta^4)^2$, and the Ricci curvature tensor is given

as follows:

$$\text{Ricci} = [\text{Ricci}_{ij}] = \begin{bmatrix} \frac{-\theta^5}{2\Delta} & \frac{\theta^4}{2\Delta} & 0 & 0 & 0 \\ \frac{\theta^4}{2\Delta} & \frac{-\theta^3}{2\Delta} & 0 & 0 & 0 \\ 0 & 0 & \frac{(-\theta^5)^2}{2\Delta^2} & \frac{\theta^4\theta^5}{\Delta^2} & \frac{\theta^3\theta^5-3(\theta^4)^2}{4\Delta^2} \\ 0 & 0 & \frac{\theta^4\theta^5}{\Delta^2} & \frac{-\theta^3\theta^5+(\theta^4)^2}{2\Delta^2} & \frac{\theta^3\theta^4}{\Delta^2} \\ 0 & 0 & \frac{\theta^3\theta^5-3(\theta^4)^2}{4\Delta^2} & \frac{\theta^3\theta^4}{\Delta^2} & \frac{(-\theta^3)^2)}{2\Delta^2} \end{bmatrix}. \tag{A.4}$$

Furthermore, the Ricci scalar curvature is given as follows: $\text{R} = g^{ij}\text{Ricci}_{ij} = -9/2$.

For $m = 3$, zero-mean, MVN with the parameter space of $\theta = \{\theta^i\}_{i=1,\dots 6} = \{\sigma_{11}, \sigma_{12}, \sigma_{13}, \sigma_{22}, \sigma_{23}, \sigma_{33}\}$, the tensor components are more involved. Hence, we directly express the Ricci scalar curvature below (see [77]):

$$\begin{aligned} \text{R} = -0.03(&51(\theta^1\theta^4\theta^6)^3 - 99((\theta^2)^6(\theta^6)^3 + (\theta^3)^6(\theta^4)^3 + (\theta^5)^6(\theta^1)^3) - \\ &161((\theta^1\theta^3\theta^6)^2(\theta^4)^3 + (\theta^5\theta^4\theta^6)^2(\theta^1)^3 + (\theta^1\theta^2\theta^4)^2(\theta^6)^3) + \\ &209((\theta^1)^3(\theta^5)^4\theta^4\theta^6 + \theta^1(\theta^2)^4(\theta^6)^3\theta^4 + \theta^1(\theta^3)^4(\theta^4)^3\theta^6) - \\ &221((\theta^1)^2(\theta^2)^2(\theta^5)^4\theta^6 + 221\theta^1(\theta^2)^4(\theta^5)^2(\theta^6)^2 + (\theta^1)^2(\theta^3)^2\theta^4(\theta^5)^4) + \\ &221(\theta^1(\theta^3)^4(\theta^4)^2(\theta^5)^2 + (\theta^2)^2(\theta^3)^4(\theta^4)^2\theta^6 + (\theta^2)^4(\theta^3)^2(\theta^6)^2(\theta^4)) + \\ &306((\theta^1)^2\theta^2\theta^3\theta^4(\theta^5)^2(\theta^6)^2) + 426(\theta^1(\theta^2)^2(\theta^3)^3\theta^4(\theta^5)^2\theta^6) + \\ &366((\theta^1)^2(\theta)^2\theta^4(\theta^5)^2(\theta^6)^2 + \theta^1(\theta^2)^2(\theta^3)^2(\theta^4)^2(\theta^6)^2) + \\ &366((\theta^1)^2(\theta^3)^2(\theta^4)^2(\theta^5)^2\theta^6) + 768(\theta^2\theta^3\theta^5)^3 \\ &522((\theta^1)^2\theta^2\theta^3(\theta^5)^5 + (\theta^2)^5\theta^3(\theta^6)^2\theta^5 + \theta^2(\theta^3)^5\theta^5(\theta^4)^2) - \\ &796(\theta^1\theta^2(\theta^3)^3(\theta^4)^2\theta^5\theta^6 + \theta^1(\theta^2)^3\theta^3\theta^4\theta^5(\theta^6)^2 + (\theta^1)^2\theta^2\theta^3\theta^4(\theta^5)^3\theta^6) + \\ &940((\theta^2)^3(\theta^3)^3\theta^4\theta^5\theta^6 + \theta^1(\theta^2)^3\theta^3(\theta^5)^3\theta^6 + \theta^1\theta^2(\theta^3)^3\theta^4(\theta^5)^3) - \\ &1056(\theta^1(\theta^2)^2(\theta^3)^2(\theta^5)^4 + (\theta^2)^4(\theta^3)^2(\theta^5)^2\theta^6 + (\theta^2)^2(\theta^3)^4\theta^4(\theta^5)^2))/\Delta^3, \end{aligned} \tag{A.5}$$

where the determinant is $\Delta = |\Sigma| = \theta^1\theta^4\theta^6 - \theta^1(\theta^5)^2 - (\theta^2)^2(\theta^6) + 2\theta^2\theta^3\theta^5 - (\theta^3)^2\theta^4$. An alternative way to compute the identities of the Riemannian geometry is through algorithmic approximations. One such iterative algorithm is described in [147].

A.2 Distance Functions on a Statistical Manifold

The geodesic distance on a statistical manifold of SPD spaces has been addressed by several authors, including [12, 68, 69, 73, 77, 78, 91, 148, 149]. One of the first known expressions for the geodesic distance between members of the MVN family with a common mean and different variance-covariance matrix was introduced in [90]. The formulation of Equation 2.4, which we have used extensively in our geostatistical models, is a result of [90, 91]. A smooth curve on the statistical manifold of MVN family is given as follows:

$$\mathrm{dt}^2 = \mathrm{d}\mu^T \, \Sigma^{-1} \, \mathrm{d}\mu + \frac{1}{2}\mathrm{tr}\Big(\left(\Sigma^{-1} \, \mathrm{d}\Sigma\right)^2 \Big), \tag{A.6}$$

where $\mathrm{t} \in [a, b]$, and the parameters μ and Σ describe the manifold of MVN family. Then, the basic solution for the geodesic, $D(\cdot, \cdot)$, is to solve the system of differential equations for the piecewise-wise smooth curves, or line segments, $\gamma : [a, b] \to \mathcal{M}$, on a manifold, $\mathcal{M}$, that satisfy the following conditions:

$$\ddot{\mu} - \dot{\Sigma}\Sigma^{-1}\dot{\mu} = 0, \quad \ddot{\Sigma} + \dot{\mu}\dot{\mu}^T - \dot{\Sigma}\Sigma^{-1}\dot{\Sigma} = 0, \tag{A.7}$$

where $\{\dot{\cdot}\} = \frac{\mathrm{d}}{\mathrm{dt}}$. Assuming the Fisher information metric, as described in Section 2.2.3, one can use the invariance under congruence transformation property to aid the above solution of the differential equation. Further details of the proof for the geodesic on a submanifold with a common mean, $\mathcal{M}_{\mu_0} = \{\mathcal{N}(\mu_0), \Sigma) | \Sigma \in \mathrm{S}^+(m, \mathbb{R})\}$, can be found in [91, Theorem 6.7].

A straightforward extension to the geodesic distance on a submanifold, $\mathcal{M}_{\mathrm{diag}} = \{\mathcal{N}(\mu), \Sigma) | \Sigma \in \mathrm{diag}((\lambda^1)^2, \ldots, (\lambda^m)^2); \mu \in \mathbb{R}^m\}$, for the uncommon mean and diagonal variance-covariance matrix can be written as follows:

$$D_{\mathrm{diag}}(P_1, P_2) = 2 \sum_{i=1}^{m} \cosh^{-2}\Big(\frac{(\mu_1^i - \mu_2^i)^2 + 2((\lambda_1^i)^2 + (\lambda_2^i)^2)}{4\lambda_1^i \lambda_2^i} \Big), \tag{A.8}$$

where $P_i = \mathcal{N}(\mu_i, \mathrm{diag}((\lambda^1)_i^2, \ldots, (\lambda^m)_i^2))$. A more general solution of the geodesic distance for an uncommon mean MVN family, and endowed with the Fisher information

metric, is described in [65].

Some of the essential properties of this geodesic distance, which aided our geostatistical model development, is enumerated below:

1. (Positivity) $D(\Sigma_1, \Sigma_2) \geq 0$, and $D(\Sigma_1, \Sigma_2) = 0 \Leftrightarrow \Sigma_1 = \Sigma_2$
2. (Symmetry) $D(\Sigma_1, \Sigma_2) = D(\Sigma_2, \Sigma_1)$
3. (Triangle inequality) $D(\Sigma_1, \Sigma_2) + D(\Sigma_2, \Sigma_3) \geq D(\Sigma_1, \Sigma_3)$
4. (Invariance under inversion) $D(\Sigma_1, \Sigma_2) = D(\Sigma_1^{-1}, \Sigma_2^{-1})$
5. (Invariance under transformation) $D(\Sigma_1, \Sigma_2) = D(A\Sigma_1 A^T, A\Sigma_2 A^T), A \forall \mathrm{GL}(m, \mathbb{R})$

For further discussions on these properties and their proofs see [149].

A.3 Distance Functions for Spatial Processes

Many of the climate-related datasets are observed at geolocations, $\boldsymbol{x_i}$, on the surface of the Earth. The canonical geographical coordinates are denoted as follows: $\boldsymbol{x_i} = (x_i^{\mathrm{lat}}, x_i^{\mathrm{lon}})^T$. Furthermore, a canonical and accurate model of the Earth's surface is an ellipsoid. However, for spatial models, and excluding polar regions, an assumption of a spherical geometry, $\mathbb{S}^m = \{\boldsymbol{x} \in \mathbb{R}^{m+1} : m \geq 1, \|\boldsymbol{x}\| = \mathrm{rad}\}$, for the Earth's surface is sufficient.

The natural metric between two points, $\{\boldsymbol{x_i}, \boldsymbol{x_j}\} \in \mathbb{S}^m$, for this spherical geometry is the great circle distance (see [102]) and can be computed as follows: $d_{gcd} = \mathrm{rad} \triangle\theta$. Here, rad $\approx$ 6371km is the radius of the earth and is usually assumed to be constant. Furthermore, $\triangle\theta = \arccos(\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle)$ is the central angle between the geolocations, and is measured in radians, where $\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle$ is the inner product in $\mathbb{R}^m$. A computationally favorable formulation (see [150]) for the central angle is given as follows:

$$\triangle_{gcd}\theta = \operatorname{atan2}\left(\frac{\sqrt{\left(\cos(\phi_2)\sin(\Delta)\right)^2 + \left(\cos(\phi_1)\sin(\phi_2) - \sin(\phi_1)\cos(\phi_2)\cos(\Delta)\right)^2}}{\sin(\phi_1)\sin(\phi_2) + \cos(\phi_1)\cos(\phi_2)\cos(\Delta)} \right), \quad \text{(A.9)}$$

where $\Delta = |x_j^{lon} - x_i^{lon}|$, $\phi_1 = x_i^{lat}$, and $\phi_2 = x_j^{lat}$. Using the great circle distance metric for modeling processes on a sphere violates the positive definiteness of certain covariance

functions (see [151]), including the non-stationary Matérn covariance function, as given in Equation 3.12 (see [8]).

In spatial statistics, many of the popular covariance functions, including the Matérn function, are defined on the Euclidean space. Hence, in practice, using the standard Euclidean metric, $d_e = \|\boldsymbol{x} - \boldsymbol{x'}\|$, for distances less than 2000km apart usually provides a reasonable model for the local behavior of the process (see [105]). However, for far apart distances standard Euclidean distance leads to overestimation. An alternative natural metric for a valid covariance function on a sphere can be expressed as the Euclidean metric in terms of the great circle distance. This metric is also known as the chordal distance and is given as follows: $\triangle_{\text{chord}}\theta = 2\sin(\triangle_{gcd}/2)$.

For far apart distances, the chordal distance provides a better approximation for the local behavior when compared to the Euclidean distance. Nevertheless, the chordal distance still leads to underestimation (see [105]) and, in some cases, an unrealistic physical model (see [152]). Analysis of various metric arguments for the covariance functions on a sphere has been considerably studied in [105]. In our experimental setup, we found that the chordal distance provided satisfactory results for the stationary covariance function.

For the spatially-varying non-stationary covariance function, the kernel convolution argument of Equation 3.13 hinders the direct application of the chordal distance. In [8], an ad-hoc site-adaptive distance is applied. An alternative site-adaptive distance, also called the centroid-based method, which performs similar to the aforementioned chordal distance, is recommended in [105]. For our non-stationary models we apply the centroid-based method to transform the geographical coordinates to their planar projections and compute the distances using the standard Euclidean metric. The transformation is given as follows:

$$\phi_1' = \frac{\phi_1 - \bar{\phi}_1}{\max(\boldsymbol{\phi}_1) - \min(\boldsymbol{\phi}_1)}\Delta_1, \quad \phi_2' = \frac{\phi_2 - \bar{\phi}_2}{\max(\boldsymbol{\phi}_2) - \min(\boldsymbol{\phi}_2)}\Delta_2, \tag{A.10}$$

where $\{\bar{\phi}_1, \bar{\phi}_2\}$ is the centroid, or the mean, of latitude and longitude values from the

given dataset. Furthermore, $\{\Delta_1, \Delta_2\}$ is the great circle distance computed from Equation A.9 and between the points $\{(\bar{\phi}_1, \min(\boldsymbol{\phi}_2)), (\bar{\phi}_1, \max(\boldsymbol{\phi}_2))\}$ and $\{(\min(\boldsymbol{\phi}_1), \bar{\phi}_2), (\max(\boldsymbol{\phi}_1), \bar{\phi}_2)\}$, respectively. One of the disadvantages of the site-adaptive method is that new data points would require a re-computation of the projections. Hence, in large data climate monitoring systems site-adaptive distances might not be feasible. In our experiments for the non-stationary covariance functions, we found that the centroid-based method performed only marginally better than the distance procedure, as described in [8], and the standard Euclidean metric.

Appendix B

Numerical Schemes

B.1 Intrinsic Statistics on a Riemannian Manifold

B.1.1 Log Map

The Riemannian log map is numerically calculated using Algorithm 6 and is further described in [78, 82].

Algorithm 6: Computing the logarithmic map, $\mathrm{Log}_{\Sigma_1}(\Sigma_2)$, on the Riemannian manifold.

Input: Start point, Σ_1; end point, $\Sigma_2 \in S^+(m, \mathbb{R})$

Output: $\mathrm{Log}_{\Sigma_1}(\Sigma_2)$ ▷ Riemannian logarithmic map

1: $\mathrm{SVD}(\Sigma_1) \rightarrow \Gamma_1 \mathrm{D}_1 \Gamma_1^T$ ▷ Singular value decomposition

2: $\boldsymbol{g} = \Gamma_1 \sqrt{\mathrm{D}_1}$

3: $\hat{\Sigma} = \boldsymbol{g}^{-1} \Sigma_2 \boldsymbol{g}^{-T}$

4: $\mathrm{SVD}(\hat{\Sigma}) \rightarrow \Gamma_2 \mathrm{D}_2 \Gamma_2^T$

5: $\mathrm{Log}_{\Sigma_1}(\Sigma_2) \leftarrow (\boldsymbol{g}\Gamma_2) \log(\mathrm{D}_2)(\boldsymbol{g}\Gamma_2)^T$

The Riemannian exponential map, $\mathrm{Exp}_{\Sigma_1}(\Sigma_2)$, is the inverse of the Riemannian log map. Hence, using a similar diagonalization scheme as in Algorithm 6, we can compute $\mathrm{Exp}_{\Sigma_1}(\Sigma_2)$.

B.1.2 Mean and variance

The Fréchet intrinsic mean on a Riemannian manifold, as described in Equation 2.9, can be computed using the standard gradient descent minimization scheme, as outlined in Algorithm 7.

Algorithm 7: Computing the intrinsic mean, $\bar{\Sigma}$, on the Riemannian manifold.

Input: $\{\Sigma_1, \ldots, \Sigma_N\} \in S^+(m, \mathbb{R})$

Output: $\bar{\Sigma} \in S^+(m, \mathbb{R})$ ▷ Riemannian intrinsic mean

1: $\bar{\Sigma}_0 = \Sigma_1$ ▷ Initial mean estimate

2: $\gamma = 1$ ▷ Initial step size

3: **while** $\mu^2_{F_i} > \epsilon$ **do** ▷ ϵ is a very small value

4: $\quad \hat{\Sigma}_i = \frac{1}{N} \sum_{k=1}^{N} \mathrm{Log}_{\bar{\Sigma}_i}(\Sigma_k)$ ▷ $\hat{\Sigma}_i$ is the candidate of the intrinsic mean, and is computed on the tangent space at $\bar{\Sigma}_i$

5: $\quad \mu^2_{F_i} = \frac{1}{N} \sum_{k=1}^{N} \|\mathrm{Log}_{\bar{\Sigma}_i}(\Sigma_k)\|^2_F$ ▷ $\mu^2_{F_i}$ is the intrinsic mean function, which is to be minimized

6: $\quad \bar{\Sigma}_{i+1} = \mathrm{Exp}_{\bar{\Sigma}_i}(\gamma \hat{\Sigma}_i)$ ▷ $\bar{\Sigma}_{i+1}$ is the projection from the tangent space at $\bar{\Sigma}_i$ to the manifold space, and is traversed at a distance, γ, along the geodesic flow

7: $\quad$ **if** $\mu^2_{F_i} > \mu^2_{F_{i+1}}$ **then**

8: $\quad\quad \bar{\Sigma}_i = \bar{\Sigma}_{i+1}$

9: $\quad\quad \gamma = \gamma/2$ ▷ Reduce the step size

10: $\quad$ **end if**

11: **end while**

Using the intrinsic mean, as given in Algorithm 7, and the Logarithmic map, as given in Algorithm 6, we can compute the intrinsic variance, $\Lambda_{\bar{\Sigma}}$, as described in Equation 2.11.

B.2 High-Dimensional Manifold Geostatistical Models

B.2.1 High-dimensional Manifold Non-stationary Gaussian Process

Proposal Scheme 8 outlines an alternative high-dimensional manifold non-stationary Gaussian process (HD-NSGP) parameter estimation procedure. This scheme leverages an existing sampling scheme of posterior mean sampling, as described in [8], and algorithmically estimates the regional aspect of the spatially-varying geometric anisotropy for the HD-NSGP model.

Proposal Scheme 8: An alternative scheme for the spatially-varying geometric anisotropy, or scale matrix parameters, $\theta_\Sigma = \{|r|, \bar{\Sigma}, \Lambda\}$.

1: Initialize $\Sigma(\boldsymbol{x})$ from the GP processes and use the PMC sampler, as given in [8].

2: Sample $|r| \sim \mathcal{U}\{1, n\}$ either uninformatively by using a uniform discrete distribution or informatively by using the geophysics of the climate variable. Alternatively, sample $\tau_{|r|} \sim \mathcal{N}(0, 1)$ such that $\text{tr}(\Lambda_{(\Sigma_{i=1},\ldots,\Sigma_{|r|})}) \leq \tau^2_{|r|}$, where Λ is computed on the SPD manifold using Equation 2.11.

3: Represent the regionally-varying scale matrix parameter using the intrinsic mean on the SPD manifold, i.e., $\bar{\Sigma}_{r(i)} = \mu^2_F(\Sigma_{i=1}, \ldots, \Sigma_{|r|})$ as given in Equation 2.9.

B.2.2 High-dimensional Manifold Multivariate Normal Sampling

Proposal Scheme 9 outlines the high-dimensional manifold multivariate normal (HD-MVN) sampling procedure, as introduced in Chapter 6.

Proposal Scheme 9: For simulating future projections, $\tilde{\mathbf{z}}$, from climate model outputs.

1: Estimate $\{\Sigma_1, \ldots, \Sigma_n\}$ using Scheme 5.

2: Estimate clusters, N, on a manifold of covariance matrices, using the following proximity criteria: $\max\{D(\Sigma_1, \Sigma_2) : \Sigma_1 \in N_1(\Sigma_i), \Sigma_2 \in N_2(\Sigma_j)\} < \tau$, and a standard hierarchical clustering method. The choice of the threshold, τ, is vague but proper. For our experiments, we chose $\tau = 2$.

3: Estimate weighted mean, $\hat{\boldsymbol{\mu}}$, for the process, $\mathbf{Z}$, using Equation 6.1b.

4: Estimate weighted mean of the covariance matrices, $\bar{\Sigma}$, and its associated variance-covariance matrix, $\Lambda_{\bar{\Sigma}}$, using intrinsic statistics on a manifold.

5: Sample covariance matrices, $\hat{\Sigma}$, for the process, $\mathbf{Z}$, using Equation 6.1c.

6: Sample a new realization, $\tilde{\mathbf{z}}$, using Equation 6.1a.

Appendix C

Notations

Acronyms

IPCC Intergovernmental Panel for Climate Change

ESM Earth System Model

CCSM Community Climate System Model

CMIP5 Coupled Model Intercomparison Project - 5

GPCP Global Precipitation Climatological Project

PSML Permanent Service for Mean Sea Level

RCP Representative Concentration Pathways

GHG Greenhouse Gas

GIA Glacial-Isostatic Adjustment

TG Tide Gauge Stations

GMSL Global Mean Sea Level

ReSL Regional Sea-Level changes

RSL Relative Sea-Level Changes or Trends

LMSL Local Mean Sea-level Changes or Trends

HDM High-Dimensional Manifold

GP Gaussian Process (in Machine Learning) or Kriging (in Geostatistics)

PCA Principal Component Analysis

EOF Empirical Orthogonal Functions

SVD Singular Value Decomposition

IRF Intrinsic Random Function

MCMC Markov Chain Monte Carlo

MLE Maximum Likelihood Estimation

PDF Probability Density Function

MSE Mean Square Error

NLPD Negative Log Predictive Density

CRPS Continuous Rank Probability Score

STD Standard Deviation

TGP Gaussian Process with Ground Truth Parameters

KNN K-Nearest Neighbor Model

SGP Stationary Gaussian Process model

NSGP Non-Stationary Gaussian Process

HD-NSGP High-Dimensional Manifold Non-stationary Gaussian Process model

DF-NSGP Data-Fusion Manifold Non-Stationary Gaussian Process model

HD-MVN High-Dimensional Multivariate Normal Sampling Scheme

Indices and Coordinates

Ξ Set of coordinate systems

$[\xi^i(p)] = [\xi^i]$ Coordinates of a point, p, on a manifold

$\frac{\partial}{\partial \xi^i}$ Natural basis of the coordinate system $[\xi^i]$

s_i, $\boldsymbol{s_i}$ Subscript refers to the dimension in the input space, $\mathbb{R}^n$

$x^i(\boldsymbol{s})$ Superscript refers to the dimension in the covariate space, $\mathbb{R}^m$

$\{\boldsymbol{x^*}, y^*\}$ Test data points

$\{\boldsymbol{x}, y\}$ Training data points

$\tilde{y}$ New sample, or realization, of the variable y

$\bar{y}$ Mean value of the variable y

$\hat{y}$ Estimates, or samples, of the variable y

θ^i_Σ The superscript on the parameter, θ, represents an element of the parameter space at a particular level of the model hierarchy; the subscript represents the associated parameter, or process, at a level above in the hierarchy

θ_Σ A set of parameters, $\{\theta^i_\Sigma\}$, for modeling a parameter, or process, Σ, that is a level above in the model hierarchy

$\{\theta^i_\Sigma\}_{\boldsymbol{s}}$ Each parameter in the set, $\{\theta^i_\Sigma\}$, is evaluated at each of the geolocations, $\boldsymbol{s}$

θ_{est} Set of parameters to be estimated

θ_{total} Set of all of the parameters in the model

$\theta_{\text{level}=i}$ Set of parameters to be estimated at a level, i, in the model hierarchy

Operations

∇ Vector differential operator

$\langle \cdot, \cdot \rangle$ Inner product

$\oplus$ Tensor addition on a manifold

$\odot$ Tensor product on a manifold

$\text{tr}(\Lambda)$ Trace of a matrix Λ

$|\Lambda|$ Determinant of a matrix Λ

$\triangleq$ Variable or notation assignment

$\equiv$ Equivalent by definition

Variables and Parameters

$\boldsymbol{s}$ Geographic locations or geolocations, i.e., $\boldsymbol{s} = \{\boldsymbol{s_1}, \ldots, \boldsymbol{s_n}\}$

$\boldsymbol{x}$ Locations in the covariate or feature space

m Dimensions of the covariate space (in statistics) or feature space (in machine learning)

n Number of geolocations

dim Dimension of the process, or the physical space of the geolocation s

$Z(\boldsymbol{s})$ Random variable at geolocations $\boldsymbol{s}$

$Y(\boldsymbol{s})$ observation at geolocations $\boldsymbol{s}$

$\epsilon(\boldsymbol{s})$ Noise variable at geolocations $\boldsymbol{s}$

p, q Point on a manifold

lat, lon Latitude, longitude

$\boldsymbol{s^t}$ Transformed geolocation

θ Set of parameters of a function

λ Set of eigenvalues of a matrix

$\{u, v\}$ Elements of the eigenvector matrix

s^2, σ^2 Scalar variance

$\mathrm{r}_{\mathrm{inj}}$ Injectivity radius

r Range parameter of the semi-variogram model

σ_Z^2 Signal variance for the covariance function

σ_Y^2 Noise variance for the covariance function

ν Smoothness parameter of the Matérn function

l Characteristic length scale

h Interpoint distance between geolocations

ψ_A Geometrical anisotropy angle

ψ_R Geometrical anisotropy ratio

μ Mean of a distribution

σ_Σ^2 Empirical Riemannian scalar variance for Riemannian mean, Σ

k_G, k_{SG}, k_{scG} Normalization factor for f_G, f_{SG}, f_{scG}

R Ricci scalar curvature

Surr Surrogate source of the dataset

Matrix and Vectors

A Affine matrix

Σ Symmetric positive definite matrix; Contextually, Σ either represents a geometric anisotropic matrix or a covariance matrix

L Lower triangular matrix such that $LL^T = \Sigma$

K Covariance matrix

$\mathbf{1}$ Vector of ones

$\mathbf{I}$ Identity Matrix

D Diagonal matrix of eigenvalues

Γ Rotational matrix or matrix of eigenvectors

H Matrix computed using the correlation function

Q Scaled distance matrix of dimension $\boldsymbol{s} \in \mathbb{R}^n$

X Tangent vectors

Λ Standard variance-covariance matrix for MVN family of probability distributions on a statistical manifold

Λ_R Riemannian variance-covariance matrix for MVN family of probability distributions on a statistical manifold

Λ_Σ Empirical Riemannian variance-covariance matrix for Riemannian mean, Σ

Ricci Ricci curvature tensor

$\mathcal{R}$ Riemannian curvature tensor

K Sectional curvature tensor

$\boldsymbol{\beta}$ Tangent vector on a statistical manifold

Functions

$Cov(\cdot, \cdot)$ Covariance

$Var(\cdot)$ Variance

$Corr(\cdot, \cdot)$ Correlation

$E(\cdot)$ Expectation of a distribution

$V(\cdot, \cdot)$ Variogram function

$k(\cdot)$, $k(\cdot, \cdot)$ Stationary covariance function, non-stationary covariance function

$\rho(\cdot)$, $\rho(\cdot, \cdot)$ Stationary correlation function, non-stationary correlation function

$\rho^S(\cdot)$ Stationary correlation function

$\rho^{NS}(\cdot, \cdot)$ Spatially-varying non-stationary correlation function

$\rho^{HDNS}(\cdot, \cdot)$ Spatially-varying high-dimensional manifold non-stationary correlation function

$\rho^{DFNS}(\cdot,\cdot)$ Spatially-varying data-fused non-stationary correlation function

$q(\cdot,\cdot)$ Scaled distance function between s and s'

$K_\nu(\cdot)$ Modified Bessel function of second kind of order ν

$M_\nu(\cdot)$ Matérn covariance function

$\Gamma(\cdot)$ Gamma function

$\Sigma(\cdot)$ Symmetric positive definite matrix function

Σ_{ij} Geometrical anisotropy model that is a function of a pair of geolocations, i.e., $\Sigma(s_i, s_j)$

$r(i)$ A function that identifies the region on the arbitrary space associated to the location s_i in the input space

$d_{gcd}(\cdot,\cdot)$ Great circle distance

$d_e(\cdot,\cdot)$ Standard Euclidean distance

$d(\cdot,\cdot)$ Arbitrary distance function

$g(\cdot,\cdot)$ Riemannian metric

$D(\cdot,\cdot)$ or $d_g(\cdot,\cdot)$ Riemannian distance function

$\mathrm{L}(\gamma)$ Length or norm of a curve γ

$\gamma(\cdot)$ Curve on a manifold

Exp or $\exp_{\Sigma_i}(\Sigma_j)$ Exponential map from Σ_i to Σ_j

Log or $\log_{\Sigma_i}(\Sigma_j)$ Logarithmic map from Σ_i to Σ_j

$\exp(\cdot)$, $\log(\cdot)$ Exponential function, logarithmic function

$f_G(\cdot)$ Probability density function for a generalized normal distribution

$f_{SG}(\cdot)$ Probability density function for a simplified generalized normal distribution

$f_{scG}(\cdot)$ Probability density function for a generalized normal distribution with scalar variance, σ^2

$\text{erf}(\cdot)$ Error function

$\mu_F(\cdot)$ Fréchet mean function

I Indicator function

$\mathcal{L}(\cdot)$ Likelihood function

$\mathcal{O}(\cdot)$ Big oh notation, i.e., for $f(\cdot), g(\cdot) \in \mathbb{N}$, $f(n) = \mathcal{O}(g(n))$, if $\frac{f(n)}{g(n)}$ is bounded as $n \to \infty$

$\boldsymbol{x}(\boldsymbol{s})$ Coordinate transformation to the covariate space

Distributions

$\{Z(\boldsymbol{s})\}$ Stochastic process

$\text{MVN}(\cdot, \cdot)$ Multi-variate normal distribution

$\mathcal{N}(\cdot, \cdot)$ Normal distribution

$\mathcal{U}(\cdot, \cdot)$ Uniform continous distribution

$\mathcal{U}\{\cdot, \cdot\}$ Uniform discrete distribution

$\mathcal{IG}(\cdot, \cdot)$ Inverse gamma distribution

$\text{GP}(\cdot, \cdot)$ Gaussian process

$p(x; \theta)$ Probability model for x parameterized by θ

$\eta(\cdot)$ Noise distribution

$w(\cdot)$ White noise process

Spaces

$\mathcal{A}$ Set of coordinate systems

$\mathbb{R}, \mathbb{N}$ Set of real numbers, Set of natural numbers

$\mathcal{A}_1$ Affine subspace

$\mathcal{I}$ Arbitrary input space

$\mathcal{B}$ Borel field

$\mathcal{X}$ Set of probability distributions

$\{p_\xi\}$ Family of probability distributions with coordinates $[\xi]$

$\{p(\cdot|\Sigma)\}$ MVN family of probability distributions

$\mathcal{G}$ Set of points that lie on a globe

$\mathcal{M}$ Manifold of a set of points that are symmetric positive definite matrices

$\mathcal{S}$ Manifold of a set of points

$\mathcal{S}(m)$ Manifold with dimension m

SPD Symmetric positive definite space

$\mathbb{S}^m$ Sphere of dimension m

$S^+(m, \mathbb{R})$ Statistical manifold of MVN family of distributions with fixed or zero mean and variance-covariance matrix of dimension m

$\mathcal{T}_{\Sigma_i}\mathcal{M}$ Tangent space at point, Σ_i, on a manifold, $\mathcal{M}$

$\mathrm{SO}(m)$ Special orthogonal group of dimension m

$\mathrm{GL}(m, \mathbb{R})$, $\mathrm{gl}(m, \mathbb{R})$ Lie group, lie algebra of dimension m with entries in $\mathbb{R}$

$\mathrm{O}(m)$ Orthogonal group of dimension m

C^∞ Infinitely many times differentiable