SARANGAN BALASUBRAMANIAM, Ph.D.
SPATIAL PREDICTION FOR AXIALLY SYMMETRIC PROCESS ON SPHERES.
(2022).
Directed by Dr. Haimeng Zhang. 102 pp.

Spatial prediction, or so-called kriging, is one of the ultimate goals in spatial data analysis. The basic idea of kriging is to use the values of a geographic variable at some locations to estimate the value(s) that are unknown at other locations. In this dissertation, we consider the spatial prediction when a random process is axially symmetric on the sphere. More specifically, we first decompose an axially symmetric process as Fourier series on circles, where the Fourier random coefficients can be expressed as circularly-symmetric complex random processes. The estimation of covariance functions for complex random processes is then obtained through both parametric and non-parametric approaches, where least squared error estimation and the Wavelet-Galerkin methods are applied, respectively. Ordinary kriging is then conducted on possibly complex random fields and predicted data values are computed through the inverse Discrete Fourier transformation. All the above approaches and results are demonstrated through simulation studies.

SPATIAL PREDICTION FOR AXIALLY SYMMETRIC PROCESS ON SPHERES

by

Sarangan Balasubramaniam

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2022

Approved by

Haimeng Zhang
Committee Chair

To my family.

APPROVAL PAGE

This dissertation written by Sarangan Balasubramaniam has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____
Haimeng Zhang

Committee Members _____
Nicholas Bussberg

_____
Xiaoli Gao

_____
Sat Gupta

_____
Jianping Sun

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGMENTS

I know, these words cannot express my acknowledgement to my supervisor Dr. Haimeng Zhang. This is not possible without his support, encouragement and kindness. He has provided me countless advice, given me his wisdom, and shared with me his knowledge. His office door has always been open for me for questions and discussions. Moreover, I wish to thank to the members of the committee Dr. Gupta, Dr. Gao, Dr. Sun, and Dr. Bussberg for the wonderful suggestions to make improvements to this dissertation.

Furthermore, I would like to acknowledge all my teachers at UNCG. I wish to acknowledge the University of North Carolina at Greensboro, particularly, the Department of Mathematics and Statistics for providing me financial support. I wish to thank Dr. Romesh, Dr. Nalin for the helpful guidance since I joined at UNCG.

My family is always my motivation for study and work. My parents are always there to support me no matter what. I would like to convey my heartfelt thanks to my parents, my brother Sethusanker (Bhanuja), my sisters Theepan Piramila, Nirusha, my niece Shaziha and my nephews Sharveen and Avaneesh for bring endless happiness and smile to the family.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

1.1 Introduction to Spatial Statistics

Spatial statistics refers to the application of statistical concepts and methods to data that have spatial location, and this location component is used as a pivotal and essential part of analysis [UNWIN09]. Using statistical techniques, topological, geometric, or geographic properties of data are analyzed and inferences are deduced. The GIS dictionary [Wad06] defines spatial statistics as the field of study in which statistical methods are used to do computations on space and spatial relationships such as distance, length, height, orientation centrality and other related characteristics. Spatial analysis began with early attempts of land surveying and had insubstantial growth until the early 1990‘s. In the last four decades with the contribution from other branches of science, spatial statistics has witnessed a quantum leap in computational science. Many fields of study had contributed to the growth of spatial statistics. Biology through botanical studies, economics through spatial econometrics and epidemiology using disease mapping helped the evolution of spatial statistics in the initial phase. Currently many studies including image processing, mineral environmental, earth science, ecology, climatology, real estate marketing, remote sensing, agronomy are contributing to the research, academic literature and application of spatial statistics.

1.1.1 Spatial Data

The data used in spatial statistics is called spatial data. Spatial data is any type of data that directly or indirectly references a specific geographical area or location. There are a variety of popular spatial data sets available in spatial statistics research, including but not limited to the Total Ozone Mapping Spectrometer (TOMS) [Ste07], Large ENSemble (LENS), and Microwave Sounding Units (MSU). The following figure (Figure 1) depicts the spatial data which carried a TOMS instrument that measured total column ozone daily from November 1, 1978 to May 6, 1993. A scanning mirror repeatedly scanned across track about 3000km wide to collect the data on each track yielding 35 total column ozone measurements.

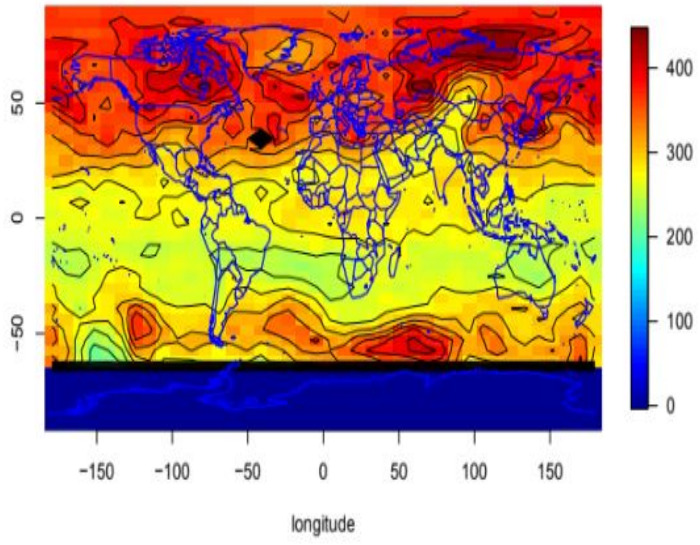

Figure 1. TOMS data; data resolution spatial 1^0 Latitude $\times$ 1.25^0 Longitudes in May, 1 - 6, 1990 [Van16].

1.1.2 Spatial Data Analysis

Spatial data analysis provides a set of tools used to analyze the spatial data. [Good08], describes spatial data analysis as a set of techniques used to find patterns, detect anomalies or test hypotheses and theories, centered on spatial data. More precisely, the technique used in the study can be called spatial analysis, if and only if the result obtained is unchanged even if the object under analysis undergoes relocation. Scrupulously, the location of the object under study is the base for the analysis. [BISHOP20] inferred that spatial data analysis is guided by spatial concepts where mathematics, geo-statistics and myriad of other techniques can be applied to deduce a pattern, characterize a phenomenon and make inferences.

1.1.3 Spatial Prediction

As per [Zhu18], the main aim of spatial prediction is to use the already known information about location to estimate or predict the values of a geographic variable (referred to as the target variable) which are unknown in the same or other locations. Spatial prediction is one of the fastest growing research topics in spatial data analysis. A variety of approaches and techniques are used in spatial statistics to make prediction. Unlike, other prediction models in statistics, spatial prediction is quite complex and challenging due to the addition of location information to the data. Kriging is one of the popular methods used in predicting data values on locations in a spatial field. Universal kriging and ordinary kriging are the most commonly used kriging methods. In this dissertation, we will focus on ordinary kriging.

The following graph illustrates the challenges of spatial prediction. Here we have 10 observed values on points $s_1, \cdots, s_{10}$ (observed data) and one unknown value at the point s_0 (unobserved). Our goal here is to predict the value at the point s_0. With only 10 values at locations $s_1, \cdots, s_{10}$, we need to establish relationships among the observed values and the predicted value based on their geographical locations. Such relationships are often characterized by a covariance function, or in this case, by a covariance matrix of 10 by 10 for these 10 values (more specifically, 10 random variables). This will become extremely difficult if the number of geo-locations is large. Further assumptions about the underlying process, in particular, the dependency of observed and unobserved data, are needed, as they become critical in making inferences and performing kriging in spatial statistics.

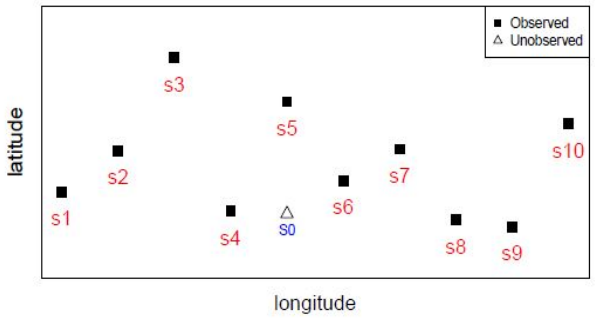

Figure 2. Arbitrary spatial data at 10 locations to demonstrate the basic process of spatial analysis

1.2 Random Process, Stationarity, and Covariance Function

In this section, we focus on covariance functions, variogram functions and various types of staitonarities for random processes on R^d, a d-dimensional Euclidean space.

1.2.1 Random Process

A random process, also called a stochastic process, is a collection of random variables $\{Z(t)|t \in T\}$ indexed by a set T. T could be a subset of real numbers, $\{1, 2, 3, \cdots\}$, $[0, \infty)$, R^d, or S^2 (a three dimensional sphere). For such a random process $Z(t)$ on a region $T \subseteq \mathsf{R}^d$, we define the finite-dimensional distributions of $Z(t)$ as the the joint distributions of $(Z_{t_1}, Z_{t_2}, \cdots, Z_{t_n})$, for $t_1, t_2, \cdots, t_n \in T$, and for any $n \in \mathsf{N}$, a set that contains all positive integers. Mathematically, the distribution of $Z(t)$ is uniquely determined by its finite dimensional distributions, according to Daniell-Kolmogorov existence theorem. More explicitly, the Daniell-Kolmogorov theorem states that to specify a stochastic process, the joint distributions of any finite subset $\{Z(t_1), \cdots, Z(t_n)\}$ must be given in a consistent way that requires:

$$P(Z(t_i) \in A_i, i = 1, \cdots, n, Z(t_{n+1}) \in R) = P(Z(t_i) \in A_i, i = 1, \cdots, n);$$
$$P(Z(t_{r_i}) \in A_{r_i}, i = 1, \cdots, n) = P(Z(t_i) \in A_i, i = 1, \cdots, n),$$

where $\pi = (r_1, r_2, \cdots, r_n)$ is a permutation of $\{1, 2, \cdots, n\}$. In other words, Kolmogorov existence theorem states that the stochastic process model is valid if the family of the finite-dimensional joint distributions is consistent under reordering of the sites and marginalization.

1.2.2 Covariance Function and Stationarity

The other important concept is the covariance. In statistics, covariance is a measure of how much two variables change together, and the covariance function describes the spatial or temporal relationship of a random process or random field. For a random field or stochastic process $Z(t)$, a covariance function $C(t, s)$ gives the covariance of the values of the random field at the two locations t and s.

Stationarity is a statistical property of a random process generating a time series that does not itself change over time. Intuitively speaking, the time series from a random process resembles itself over time intervals. Mathematically, this assumption holds when

$$E[Z(s)] = \mu; \quad Cov[Z(s), Z(t)] = Cov[Z(s+h), Z(t+h)],$$

for all shifts h.

1.2.2.1 Strong Stationarity

Given the random function $Z(t)$, strong stationarity, also referred as strict stationarity, indicates that for any number n of any sites $t_1, t_2, \cdots, t_n$, the joint cumulative function of $(Z(t_1), Z(t_2), \cdots, Z(t_k))$ remains the same under an arbitrary translation h:

$$P(Z(t_1), Z(t_2), \cdots, Z(t_k)) = P(Z(t_1+h), Z(t_2+h), \cdots, Z(t_k+h)).$$

If the process is strongly stationary and has a finite second moment, it is also weakly stationary.

1.2.2.2 Weak Stationarity

A weaker form of stationarity is the assumption that a covariance function, $Cov[Z(t+h), Z(t)]$ exists and depends only on h (equivalent to stationarity of the mean and variance only). This form of stationarity is known as second order stationarity. A process $Z(t)$ is said to be weakly stationary when,

- The mean of the process does not depend on $t : E[Z(t)] = \mu$,
- The variance of the process does not depend on $t : E[(Z(t) - \mu)^2] = \sigma^2 < \infty$,
- The covariance between $Z(t)$ and $Z(t + h)$ only depends on h,

$$\text{Cov}[Z(t), Z(t + h)] = E[(Z(t) - \mu)(Z(t + h) - \mu)] = C(h).$$

1.2.2.3 Intrinsic Stationarity

An even weaker form of the stationarity assumption was proposed by [Math71]. If the assumption of stationarity is made not about the random function $Z(t)$ but about the first order differences, $Z(t + h) - Z(t)$, then this form of stationarity is known as the intrinsic hypothesis.

Under the intrinsic hypothesis, $Z(t)$ is (intrinsically) stationary if

- the mean of the process does not depend on $t : E[Z(t)] = \mu$,
- $E[Z(t + h) - Z(t)] = 0$, for all t and h,

- $\text{Var}[Z(t+h) - Z(t)] = E\{[Z(t+h) - Z(t)]^2\} = 2\gamma(h)$, for all t and depends only on h where $\gamma(h)$ denotes the semi-variogram.

Under intrinsic stationarity, variogram or semivariogram functions play a critical role in characterizing the properties of a process. The variogram is defined as the variance of the difference between field values at two locations (t, s) across realizations of the field [Cre93]. The quantity $2\gamma(\cdot)$ is called the variogram, and $\gamma(\cdot)$ is the semivariogram. [Math71] proposed the use of the semivariogram or variogram, γ, which he defined as $\gamma(h) = \frac{1}{2}\text{Var}(\gamma(s+h) - \gamma(s))$, as an alternative to the covariance function.

A variogram function has the following properties:

(1) $\gamma(h) = \gamma(-h), \gamma(h) \geq 0$ and $\gamma(0) = 0$.

(2) Variograms are conditionally negative definite (c.n.d.): $\forall a_1, a_2, \cdots, a_n \in \mathsf{R}$ s.t. $\sum_{i=1}^{n} a_i = 0$, $\forall \{s_1, s_2, \cdots, s_n\} \subseteq T$, we have:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \gamma(s_i - s_j) \leq 0.$$

(3) If the process $Z(t)$ is stationary with covariance function $C(\theta)$, one can always find a valid semivariogram $\gamma(\theta) = C(0) - C(\theta)$.

(4) If γ is bounded in a neighborhood of 0, $\exists a$ and $b \geq 0$ such that for any $x \in \mathsf{R}^d$, $\gamma(x) \leq a||x||^2 + b$.

1.2.3 Relationships among types of Stationarity

We now establish a few propositions.

Proposition 1.1: If the process $\{Z(t); t \in \mathsf{Z}\}$ is strongly stationary and has finite second moment, then $\{Z(t); t \in \mathsf{Z}\}$ is weakly stationary.

Proof : If the process $\{Z(t); t \in \mathsf{Z}\}$ is strongly stationary, then $\cdots, Z(-1), Z(0), Z(1), \cdots$ have the same distribution function and $(Z(t_1), Z(t_2))$ and $(Z(t_1+h), Z(t_2+h))$ have the same joint distribution function for all t_1, t_2 and h. Because, by assumption, the process $\{Z(t); t \in \mathsf{Z}\}$ has a finite second moment, this implies that

- $E(Z(t)) = \mu, \quad \forall t;$
- $\mathrm{Cov}(Z(t_1), Z(t_2)) = \mathrm{Cov}(Z(t_1+h), Z(t_2+h)), \quad \forall t_1, t_2, h,$

concluding the proof.

Remark 1: If the existence of finite second moments are not assumed in the proposition, strong stationarity does not necessarily imply weak stationarity. For example, an independent and identically distributed (*i.i.d.* for short) process with standard Cauchy distribution is strictly stationary but not weakly stationary because the second moment of the process may not finite, neither the covariance function.

Remark 2: A weakly stationary process is not necessarily strongly stationary. Consider the following example. Let $\{Y(t); t \in \mathsf{Z}\}$ be a stochastic process defined by

$$Y(t) = \begin{cases} V(t), & \text{if } t \text{ is even;} \\ \frac{1}{3}(2V(t)^2 - 1), & \text{if } t \text{ is odd,} \end{cases}$$

where $V(t) \sim N(0,1)$. This process is weakly stationary but it is not strictly stationary since we have

$$E(Y(t)) = \begin{cases} E(V(t)) = 0, & \text{if } t \text{ is even;} \\ \frac{1}{\sqrt{3}} E[(2V(t)^2 - 1)] = 0, & \text{if } t \text{ is odd,} \end{cases}$$

and

$$Var(Y(t)) = \begin{cases} Var(V(t)) = 1, & \text{if } t \text{ is even;} \\ \frac{1}{3} Var[(2V(t)^2 - 1)] = 1, & \text{if } t \text{ is odd.} \end{cases}$$

Further, because $Y(t)$ and $Y(t-k)$ are independent random variables, we have $\text{Cov}(Y(t), Y(t-k)) = 0 \ \forall k$. Thus, the process $Y(t)$ is weakly stationary. In particular, it is the so-called White Noise process, denoted as $Y(t) \sim WN(0,1)$ with mean zero and variance of one.

It must be noted that $P(Y(t) \leq 0) = P(V(t) \leq 0) = 0.5$ for t is even. and $P(Y(t) \leq 0) = P(\frac{1}{\sqrt{3}}(2V(t)^2 - 1) \leq 0) = P(2V(t)^2 - 1 \leq 0) = P(V(t)^2 \leq \frac{1}{\sqrt{2}}) = 0.6$ for t is odd. Hence the random variables of the process are not identically distributed. This implies that the process is not strongly stationary.

Remark 3: There is one important case in which weak stationarity implies strong stationary.

Proposition 1.2: If $\{Z(t); t \in \mathsf{Z}\}$ is a weakly stationary Gaussian stochastic process, then $\{Z(t); t \in \mathsf{Z}\}$ is strongly stationary.

Proof: Let $\{Z(t); t \in \mathsf{Z}\}$ be a Gaussian stochastic process. Assume that the process is weakly stationary. If the process is weakly stationary, then

- $E(Z(t)) = \mu, \quad \forall t;$
- $\mathrm{Cov}(Z(t_1), Z(t_2)) = \mathrm{Cov}(Z(t_1 + h), Z(t_2 + h)), \quad \forall t_1, t_2, h,$

In addition, the joint distribution of $(Z(t_1), Z(t_2), \cdots, Z(t_s))$ depends only on the mean and covariance function. Hence, it follows that the joint distribution function of the vector $(Z(t_1), Z(t_2), \cdots, Z(t_s))$ is equal with one of $(Z(t_1 + h), Z(t_2 + h), \cdots, Z(t_s + h))$ for any finite set of indices $\{t_1, t_2, \cdots, t_s\} \subset \mathsf{Z}$ with $s \in \mathsf{Z}^+$, and $h \in \mathsf{Z}$. This implies that the process $\{Z(t); t \in \mathsf{Z}\}$ is strongly stationary.

Now we explore the relationship between weak stationarity and intrinsic stationarity. From the definitions, it is obvious that a weekly stationary process is an intrinsic stationary process. However, an intrinsic stationary process is not always a weakly stationary process. Here we want to show that Brownian motion is intrinsically stationary but not weakly stationary.

Brownian motion $B(t), t \geq 0$ satisfies the following:

- $B(0) = 0,$
- $E[B(t)] = 0$ and $Var(B(t)) = t,$
- $B(t)$ has independent increments,

- For $t_2 > t_1, B(t_2) - B(t_1) \sim N(0, (t_2 - t_1))$.

$B(t)$ having independent increments means that for all times $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$ the increments $B(t_n) - B(t_{n\ 1}), B(t_{n\ 1}) - B(t_{n\ 2}), \cdots, B(t_2) - B(t_1), B(t_1)$ are independent random variables.

$$
\begin{aligned}
&E(B(t_2) - B(t_1)) = E(B(t_2)) - E(B(t_1)) = 0, \\
&\mathrm{Cov}(B(t_2), B(t_1)) = \mathrm{Cov}(B(t_1), B(t_1)) + \mathrm{Cov}(B(t_2) - B(t_1), B(t_1)) \\
&= t_1 = \min(t_1, t_2).
\end{aligned}
$$

Here $\mathrm{Cov}(B(t_2), B(t_1))$ does not depend on the difference. Therefore, Brownian motion is not weakly stationary. Nevertheless, $\mathrm{Var}(B(t_2) - B(t_1))$ is a function on the difference $(t_2 - t_1)$. So, Brownian motion is intrinsically stationary, but not (weakly) stationary.

Finally, we give some examples to illustrate the above stationarity.

An *i.i.d.* process is a strongly stationary process. This follows almost immediately from the definition. Since the random variables $Z(t_1 + h), Z(t_2 + h), \cdots, Z(t_s + h)$ are *i.i.d*, we have that

$$
F_{t_1+h, t_2+h, \quad, t_s+h}(b_1, b_2, \cdots, b_s) = F(b_1)F(b_2)\cdots F(b_s).
$$

On the other hand, we also have $Z(t_1), Z(t_2), \cdots, Z(t_s)$ are ***i.i.d.*** and hence

$$F_{t_1,t_2,\quad,t_s}(b_1, b_2, \cdots, b_s) = F(b_1)F(b_2)\cdots F(b_s).$$

It can be concluded that, for any $h \in R^d$,

$$F_{t_1+h,t_2+h,\quad,t_s+h}(b_1, b_2, \cdots, b_s) = F_{t_1,t_2,\quad,t_s}(b_1, b_2, \cdots, b_s).$$

Consider the discrete stochastic process $\{Z(t); t \in \mathsf{N}\}$ where $Z(t) = A$, with $A \sim U(3, 7)$ (A is uniformly distributed on the interval $[3, 7]$). This process is strongly stationary. Whereas the discrete stochastic process $\{Z(t); t \in \mathsf{N}\}$ where $Z(t) = tA$, with $A \sim U(3, 7)$ is not strongly stationary.

An example of a weakly stationary process is the white noise process. A stochastic process $\{W(t); t \in \mathsf{Z}\}$ in which the random variables $W(t); t = 0, \pm1, \pm2, \cdots$ are such that

- $E(W(t)) = 0, \quad \forall t;$
- $\mathrm{Var}(W(t)) = \sigma_w^2 < 1, \quad \forall t;$
- $\mathrm{Cov}(W(t), W(t+k)) = 0 \quad \forall t \ \forall k \neq 0.$

It is called white noise with mean 0 and variance σ_w^2, written as $W(t) \sim WN(0, \sigma_w^2)$. The first condition establishes that the expectation is always constant and equal to zero. The second condition establishes that variance is constant. The third condition establishes that the variables of the process are uncorrelated for all lags. Obviously,

$W(t)$ is (weakly) stationary.

An important example of non-stationary stochastic processes is the following. Let $\{Y(t); t \in \mathsf{N}\}$ be a stochastic process where $Y(0) = \delta < \infty$ and $Y(t) = Y(t-1)+W(t)$ for $t = 1, 2, \cdots$ with $W(t) \sim WN(0, \sigma_w^2)$. This process is called random walk. The mean of $Y(t)$ is given by

$$E(Y(t)) = \delta$$

and its variance is

$$\mathrm{Var}(Y(t)) = t\sigma_w^2.$$

Thus a random walk is not a weakly stationary process, but it is intrinsically stationary since $Var(Y(t) - Y(s)) = |t - s|\sigma_w^2$.

1.2.4 Commonly Used Covariance Functions

In this section, we introduce a number of commonly used covariance functions, including spherical covariance function, exponential covariance function, Matern covariance function, and Gaussian covariance function.

Spherical Model Covariance Function : According to [Wack03], a commonly used spectral density covariance function is the spherical model

$$C_{\text{spherical}}(h) = \begin{cases} b(1 - \frac{3jhj}{2a} + \frac{(jhj)^3}{2a^3}), & \text{for } 0 \leq \mid h \mid \leq a, \\ 0, & \text{for } \mid h \mid > a. \end{cases}$$

The parameter a indicates the range of the spherical covariance: the covariance vanishes when the range is reached. The parameter b represents the maximal value of the covariance: the spherical covariance steadily decreases, starting from the maximum b at the origin, until it vanishes when the range is reached.

Exponential Covariance Function: Based on [Wack03], the exponential covariance function model falls off exponentially with increasing distance

$$C_{\text{exp}}(h) = b \exp(-\frac{|h|}{a}),$$

where $a, b > 0$ and a determines how quickly the covariance falls off. The exponential model is continuous but not differentiable at the origin. It drops asymptotically towards zero for $|h| \to \infty$.

Matern Model Covariance Function: According to the [RW05], the Matern class of covariance functions is given by

$$C_{\text{Matern}}(h) = \frac{2^{1-v}}{\Gamma(v)}(\frac{\sqrt{2v|h|}}{l})^v K_v(\frac{\sqrt{2v|h|}}{l}),$$

with positive parameters v and l, where K_v is a modified Bessel function.

The γ-exponential Covariance Function: The γ-exponential covariance function has a similar number of parameters to the Matern class covariance function. Based on [RW05], the γ-exponential family covariance function can be written as

$$C(h) = \exp(-(|h|/l)^{\gamma}) \qquad 0 < \gamma \leq 2.$$

Gaussian Covariance Function: A Gaussian process is a collection of random variables, any finite number of which have consistent Gaussian distributions. The mean of the Gaussian process can take any value and the covariance matrix should be positive definite. Based on [RW05], if we assume $\vec{X} = (X_1, \cdots, X_k)^T$ is multivariate Gaussian with its mean function $m(\vec{x})$ and a covariance function $C(\vec{x}, \vec{y}) = \sigma^2 \exp(-\frac{1}{2l^2}(\vec{x} - \vec{y})^2)$, where $\sigma > 0$ and parameter l defining the characteristic length-scale. Moreover, the stationary Gaussian covariance function is given by

$$C(h) = \sigma^2 \exp\left\{-\frac{|h|^2}{2l^2}\right\}.$$

1.3 Introduction to Axially Symmetric Process

Following the setup discussed in [HZR12], we consider a possibly complex-valued random process $X(P)$ on a unit sphere S^2, where $P = (\phi, \lambda) \in \mathsf{S}^2$ with longitude $\lambda \in [0, 2\pi]$ and latitude $\phi \in [0, \pi]$. Assume the process is continuous in quadratic mean with respect to the location P, and has finite second moment, then $X(P)$ can be represented by spherical harmonics, with convergence of the series in quadratic

mean

$$X(P) = X(\phi, \lambda) = \sum_{v=0}^{1} \sum_{m=}^{v} Z_{v,m} e^{im\lambda} P_v^m(\cos\phi),$$

where $P_v^m(\cdot)$ is a normalized associated Legendre polynomial so that its squared integral on [-1, 1] is 1, and $Z_{v,m}$ are the coefficients satisfying

$$Z_{v,m} = \int_{S^2} X(P) e^{\ im\lambda} P_v^m(\cos\phi) d\phi.$$

Without loss of generality, the process is assumed to have mean zero, i.e., $E(X(P)) = 0$, which implies $E(Z_{v,m}) = 0$. Then, the covariance function of the process at two locations $P = (\phi_p, \lambda_p)$ and $Q = (\phi_Q, \lambda_Q)$ is given by,

$$R(P,Q) = E(X(P)\bar{X}(Q)) = \sum_{v=0}^{1} \sum_{m=}^{v} \sum_{\mu=0}^{1} \sum_{n=}^{\mu} Z_{v,m} \bar{Z}_{\mu,n} e^{im\lambda_p} P_v^m(\cos\phi_p) e^{in\lambda_q} P_v^m(\cos\phi_Q),$$

where $\bar{Z}$ denotes the complex conjugate of Z. Under the assumption of axial symmetry [Jon63], where the covariance function depends on the longitudes only through their difference, one has

$$E(Z_{v,m}\bar{Z}_{\mu,n}) = \delta_{n,m} f_{v,\mu,m}$$

where $\delta_{n,m} = 0$ if $n \neq m$ else 1.

Moreover, for an axially symmetric process $X(P), P \in \mathsf{S}^2$, the covariance function $R(P,Q)$ at two locations $P = (\phi_P, \lambda_P), Q = (\phi_Q, \lambda_Q) \in \mathsf{S}^2$ is given by

$$R(\phi_P, \phi_Q, \lambda_P, \lambda_Q) = R(\phi_P, \phi_Q, \lambda_P - \lambda_Q).$$

That is, the covariance function on two locations depend on the (directional) longitudinal difference.

Using spherical harmonics, the covariance function has the property

$$R(\phi_P, \phi_Q, \lambda_P - \lambda_Q) = R(\phi_Q, \phi_P, \lambda_Q - \lambda_P).$$

One special case of the axially symmetric process is the so-called longitudinally reversible process where the covariance function $R(P,Q)$ satisfies

$$R(\phi_P, \phi_Q, \lambda_P - \lambda_Q) = R(\phi_P, \phi_Q, \lambda_Q - \lambda_P).$$

This indicates that the covariance function between two locations depends on the (un-directional) longitudinal difference.

According to [HZR12], the covariance function for an axially symmetric process can be written

$$R(P,Q) = \sum_{m=\ 1}^{1} C_m(\phi_P, \phi_Q) e^{im\Delta\lambda}, \tag{1.1}$$

where $C_m(\phi_P, \phi_Q)$ is Hermitian and positive definite with $\sum_{m=\ 1}^{1} |C_m(\phi_P, \phi_Q)| < \infty$, and $\Delta\lambda$ can be specified as

$$\Delta\lambda = \begin{cases} \lambda_P - \lambda_Q + 2\pi, & \lambda_P - \lambda_Q \leq -\pi \\ \lambda_P - \lambda_Q, & \lambda_P - \lambda_Q \in (-\pi, \pi] \\ \lambda_P - \lambda_Q - 2\pi, & \lambda_P - \lambda_Q > \pi. \end{cases}$$

Equation (1.1) is actually the Fourier transform between $R(P, Q)$ and $C_m(\phi_P, \phi_Q)$. Therefore, through the inverse Fourier transform, $C_m(\phi_p, \phi_q)$ can be obtained by

$$C_m(\phi_P, \phi_Q) = \frac{1}{2\pi} \int_{\pi}^{\pi} R(P, Q) e^{\ im\Delta\lambda} d\Delta\lambda.$$

The derivation of $C_m(\phi_p, \phi_q)$ based on the discrete Fourier transform will be used later for kriging.

The longitudinal reversibility yields $C_{\ m}(\phi_P, \phi_Q) = C_m(\phi_P, \phi_Q)$, which leads to the following representation mentioned in [HZR12] Proposition 3.

$$R(\phi_P, \phi_Q, \Delta\lambda) = C_0(\phi_P, \phi_Q) + \sum_{m=1}^{1} C_m(\phi_P, \phi_Q)(e^{\ im\Delta\lambda} + e^{im\Delta\lambda}) = \sum_{m=0}^{1} C_m(\phi_P, \phi_Q) \cos m\Delta\lambda$$

where $\Delta\lambda = \min\{|\lambda_P - \lambda_Q|, 2\pi - |\lambda_P - \lambda_Q|\} \in [0, \pi]$, and the last C_m is rescaled to absorb a constant value 2 for $m > 0$.

CHAPTER II

LITERATURE REVIEW

The word "kriging" in the domain of spatial statistics is widely used to refer to the spatial prediction methods, named in honor of D. G. Krige, a professor at the University of the Witwatersrand, South Africa. Professor Krige promoted the use of statistical tools for exploring minerals and in 1951 laid the foundation for the development of spatial statistics, which was later adopted and developed by Georges Matheron and colleagues at L'Ecole des Mines in Fontainbleau. [Ord83] describes kriging as the concept of interpolation of random spatial processes and presented predictors that were linear in observations and mentioned nonlinear possibilities. [Math63] defines kriging as predicting the grade of a panel by computing the weighted average of available sample. The suitable weights a_i is determined by $\sum_{i=1}^{n} a_i = 1$ and the variance used for prediction should take the smallest value. [Krig] perceives the word kriging as the multiple regression procedure to arrive at the best linear unbiased predictor or best linear weighted moving average predictor of the ore grade of an ore block of any size by assigning an optimum set of weights to all the relevant and available data both inside and outside of the ore block. Thus, originally geostatistical methods were developed to predict the likely yield of a mining operation over a spatial region D, from the available samples of ore extracted from a finite set of locations.

There are various kriging methods in literature. [Cre93] explains ordinary kriging as the linear prediction of points at unobserved sites. Their discussion on ordinary

kriging considers n observations $\{Z_1, \cdots, Z_n\}$ at known spatial locations $\{t_1, \cdots, t_n\}$; then the aim of kriging is to give a linear prediction Z_0 at the unobserved location t_0. The assumption is that the covariance (or variogram) of $Z(t)$ is already known. If not known it is pre-estimated. $Z(t)$ follows a Gaussian process with an unknown mean μ and constant variance. Moreover, the sum of the coefficients of a linear predictor is equal to 1. Simple kriging is another type of kriging technique similar to ordinary kriging, under which, in contrast, $Z(t)$ follows a Gaussian process with a known mean μ and the sum of the coefficients of the linear predictor is not equal to 1.

Finally, universal kriging generalizes the ordinary kriging procedure. It is kriging with a local trend. This local trend or drift varies slowly but continuously on the surface on top of which the variation to be found is superimposed. The theory behind universal kriging is to perform kriging under universality conditions, which would eventually help deepen the understanding and interpolate the points. [Wack03] states that "the universal kriging model splits the random function into a linear combination of deterministic functions, known at any point of the region, and a random component, the residual random function."

Kriging methods have been widely used in literature. [Cre88] considered the assumptions needed to carry out the spatial prediction using ordinary kriging, and analyzed how nugget effect, range, and sill of the variogram affected the predictor. Moreover, kriging is used in many fields in practical applications such as empirical Bayesian kriging (EBK) for a fast and reliable solution for both automatic and interactive data interpolation [Grib20], feed-forward neural network (FFNN)-based path loss

modeling improves the accuracy of kriging [Sato19], integrated random forest (RF) models and spatiotemporal kriging [Shao20], minimax approach used in the semivariogram fitting process [Seti20], predicting the above ground biomass map [Li20], calculating distance between locations on earth's surface [Grib20], trend removal methods for observed spatial data represented as LR - fuzzy numbers [Hes20], prediction performance of kriging predictors with isotropic Matérn correlations [Tuo20], network path prediction into statistical prediction [Chu07], novel neural network structure for spatial prediction [Li20], mapping the health risk factors in primary sector to overcome the challenges of working with discrete occupational data [Ger20], to mention just a few.

Although kriging methods have been widely used, the approaches and applications are mainly on R^d or a region of R^d. In recent years, with the global networks and satellite sensors that have been widely used to monitor a wide array of global-scale processes and variables, increasing attention has been now switched to the study of random processes on the sphere. Consequently, a number of studies on kriging methods on spheres have been proposed. For example, [Woj18] discusses the advantages and disadvantages among the widely available kriging models such as least-squares collocation, ordinary kriging, and universal kriging when real-world data has some imperfect conditions, which refer to the addition of mean value or trend to the stationary or intrinsically stationary processes on the sphere. These comparisons are carried out through the modeling of vertical total electron content (VTEC) derived from GNSS station data. [BUS20] develops a universal kriging method on the sphere based on the intrinsic random functions. They explore graphical and numerical pro-

cedures to determine the order of non-stationarity, and their proposed method has been used to analyze a multi-decadal global temperature dataset. [Grib20] provides a review on various geostatistical modeling approaches on the sphere, which include the use of covariance models that are valid in R^d, the possible new classes of covariance models that are valid on the sphere, and the use of intrinsic random functions that are proposed by [HZRS19]. They point out the limitations in the above approaches in terms of the multivariate Gaussian data assumption as well as the large number of data extents for prediction. Furthermore, they propose ways of interpolating data in a geographic coordinate system using empirical Bayesian kriging and investigate some geostatistical modeling on the sphere such as using the covariance models that allow replacement of the Euclidean distance with the great circle distance. They propose new classes of covariance functions that is not valid for Euclidean distance, while using intrinsic random functions that allows the removal of the local non-stationarity in the data.

It has been known that large scale or global processes usually show the pattern of nonstationarity since the factors driving the characteristics of the random field typically vary at various locations. In recent years, the axially symmetric process has received increased attention. For an axially symmetric processes, as described in [Jon63], the covariance function depends on the longitudes only through their difference. [Ste07] has applied this approach to model total column ozone on a global scale, while [JS08] consider the axially symmetric process by applying the differential operators to an isotropic process. [HZR12] obtained a simplified representation of a valid axially symmetric covariance function on the sphere. In this dissertation, we

focus on the ordinary kriging for the axially symmetric process on the sphere.

[Jae2014] proposed a method to construct covariance function for smooth processes on the surface of a sphere that is valid with great circle distance. Moreover, [Jae2014] used the kriging approach to numerically compare their proposed model with existing Matérn class models. Unlike our approach, instead of considering the whole sphere identically, they picked more samples near to the equator and fewer samples in both poles in a simulation method. In the comparison study, they used a cross-validation technique to justify their prediction performance. Meanwhile, they applied the maximum likelihood estimation for the parametric approach for which we had computational difficulties when we applied it to our approach. Nevertheless, the performance of the model proposed has similarities to the existing Matérn class models. However, no detail kriging approach was discussed in their paper. In [Zhu16], they proposed a kernel convolution approach to model the axially symmetric processes on the sphere. More specifically, the spatial random process is first modeled at each latitude and the raw estimates are smoothed with a local smoothing method across different latitudes through leave-one-out cross validation. They then use the estimated covariance model to perform the ordinary kriging. Finally, this approach is applied to the detrended TOMS data for 3600 locations for the ordinary kriging predictions with the assumption that the kriging variance is constant at the same latitude and the process is axially symmetric.

In this dissertation, we consider ordinary kriging when a random process on the sphere is axially symmetric. In particular, the kriging process will involve the pre-

diction for a complex random field. It would extend the ordinary kriging approach that was described in [Cre93], which focuses on random processes on the real domain. It should be noted that although many results for complex random processes can be obtained through those for real random processes, there are some differences. For example, the covariance function for complex random processes might also be complex, involving the real and imaginary components. In addition, there is very limited research on kriging under a complex domain. [Deu98] seems to be the first one to consider the complex kriging prediction, but with an additional predictor assumption. [Sand15] extensively discusses the computational developments of kriging with the complex covariance functions and performs kriging on the complex random field. Meanwhile, they applied the parametric method for estimating the parameters of the complex covariance model through the non-linear weighted least square method (for example, see [Cre93]). They applied their method to two types of data sets: regularly and irregularly sampled data.

In our dissertation research, the Discrete Fourier transform plays a pivotal role. First, according to [HZR12], an axially symmetric process on the sphere can be decomposed as Fourier series on circles, where the Fourier random coefficients can be expressed as a circularly-symmetric complex random process. Ordinary kriging is then conducted for the circularly-symmetric complex random processes, and the predicted data values will be obtained through the inverse Discrete Fourier transformation. For our kriging methods, we consider both parametric and non-parametric approaches in estimating the empirical covariance models. Under the parametric approach, we choose the optimal values of the particular parameters based on the least

total squared differences of true variograms and empirical variograms. In addition, we calculate the Increase in Prediction Error (IPE) given by [Yan13] to justify the kriging performance. For the non-parametric approach, we use the Wavelet-Galerkin method to obtain the estimated covariance function. All the above approaches and results are demonstrated through simulation studies.

The dissertation is organized as follows. In Chapter III, we provide a full development of kriging methods under both real and complex random processes. In Chapter IV, we propose both parametric and nonparametric approaches to estimate the covariance function. Finally, a summary of results and future research areas are presented in Chapter V.

CHAPTER III

SPATIAL PREDICTION FOR AXIALLY SYMMETRIC PROCESSES ON THE SPHERE

Optimal prediction (kriging) or making estimations on the spatial data analysis is always challenging and interesting. However, there is very limited research on kriging, in particular, for random processes on the sphere. In this chapter, we propose various ordinary kriging approaches when the underlying process is an axially symmetric process on the sphere.

This chapter is organized as following. We introduce some basic structure of gridded data in the spatial field, followed by variogram functions, a block circulant covariance matrix for both longitudinally reversible and axially symmetric processes on the sphere, and a cross covariance and variogram estimator on the sphere in Section 3.1. Further, in Section 3.2 we review various Fourier transforms including but not limited to the discrete Fourier transform for finite data series. Finally, Section 3.3 contains the general setup for developing our kriging methods when the underlying process is longitudinally reversible or axially symmetric processes on the sphere.

3.1 Gridded Data and Estimators

A gridded data structure defines two-dimensional coordinates based on the Earth's surface. Gridded data structure is the common way to represent the spatial data. It contains latitudes and longitudes to identify exact locations on the surface of the

Earth. The quantity of latitudes and longitudes provides the data resolution or dimension. Latitude lies between 0 and π and longitude has the range of $[0, 2\pi]$. In spatial research literature, Microwave sounding units (MSU), Total Ozone Mapping Spectrometer (TOMS) and Large ENSemble (LENS) data sets have been widely used and discussed, all of which are have gridded data structure. For example, according to [PAC18], regular gridded LENS data contain 192 latitudinal points and 288 longitudinal points. It has the 1.25 common length among longitudes. Meanwhile, MSU data have an equal length difference for both latitudes and longitudes. It has 2.5 length differences, which gives 72 latitudinal points and 144 longitudinal points on the Earth. Therefore, throughout this dissertation work we will assume that the gridded data structure on the sphere is observed.

More explicitly, we assume the data observed on the sphere are represented as $\{X(\phi_i, \lambda_j) : 1 \leq i \leq n_l, 1 \leq j \leq n\}$. Here $\phi_1, \phi_2, \cdots, \phi_{n_l}$ are n_l latitudes and $\lambda_1, \lambda_2, \cdots, \lambda_n$ are n gridded longitudinal values with $\lambda_j = (j-1)\delta$ where $\delta = 2\pi/n$ for $j = 1, 2, \cdots, n$. We can also write the gridded data in the following matrix form.

$$X = \begin{pmatrix} X(\phi_1, \lambda_1) & X(\phi_1, \lambda_2) & X(\phi_1, \lambda_3) & \cdots & X(\phi_1, \lambda_n) \\ X(\phi_2, \lambda_1) & X(\phi_2, \lambda_2) & X(\phi_2, \lambda_3) & \cdots & X(\phi_2, \lambda_n) \\ X(\phi_3, \lambda_1) & X(\phi_3, \lambda_2) & X(\phi_3, \lambda_3) & \cdots & X(\phi_3, \lambda_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X(\phi_{n_l}, \lambda_1) & X(\phi_{n_l}, \lambda_2) & X(\phi_{n_l}, \lambda_3) & \cdots & X(\phi_{n_l}, \lambda_n) \end{pmatrix}. \tag{3.1}$$

As a special case where there are only two latitudes denoted as ϕ_P and ϕ_Q, all observations on latitudes ϕ_P and ϕ_Q can be written as $\{(\phi_P, \lambda_j)\}_{j=1}^n$ and $\{(\phi_Q, \lambda_j)\}_{j=1}^n$,

respectively.

In this section, we will first discuss the cross-variogram function. We then introduce the block circulant covariance matrix for axially symmetric processes and the symmetric block circulant covariance matrix with symmetric blocks for longitudinally reversible processes on the sphere. Finally under the assumption of gridded data structure, the Method of Moments (MOM) estimators for variogram and covariance function are given.

3.1.1 Variogram functions

We assume constant means μ_1 and μ_2 on latitudes ϕ_1 and ϕ_2, respectively. For two latitudes ϕ_1 and ϕ_2, the cross variogram function on those two latitudes with longitudinal difference $\Delta\lambda$ is defined as follows and is further simplified.

$$
\begin{aligned}
& 2\gamma(\phi_1, \phi_2, \Delta\lambda) = E[(X(\phi_1, \lambda + \Delta\lambda) - X(\phi_1, \lambda))(X(\phi_2, \lambda + \Delta\lambda) - X(\phi_2, \lambda))] \\
= \;& E\left[(X((\phi_1, \lambda + \Delta\lambda) - \mu_1) - (X(\phi_1, \lambda) - \mu_1))\right. \\
& \left. \times\, (X((\phi_2, \lambda + \Delta\lambda) - \mu_2) - (X(\phi_2, \lambda) - \mu_2))\right] \\
= \;& cov(X(\phi_1, \lambda + \Delta\lambda)X(\phi_2, \lambda + \Delta\lambda)) - cov(X(\phi_1, \lambda + \Delta\lambda)X(\phi_2, \lambda)) \\
& -cov(X(\phi_1, \lambda)X(\phi_2, \lambda + \Delta\lambda)) + cov(X(\phi_1, \lambda)X(\phi_2, \lambda)) \\
= \;& R(\phi_1, \phi_2, 0) - R(\phi_1, \phi_2, \Delta\lambda) - R(\phi_1, \phi_2, -\Delta\lambda) + R(\phi_1, \phi_2, 0) \\
= \;& 2R(\phi_1, \phi_2, 0) - R(\phi_1, \phi_2, \Delta\lambda) - R(\phi_1, \phi_2, -\Delta\lambda).
\end{aligned}
$$

Hence,

$$\gamma(\phi_1, \phi_2, \Delta\lambda) = R(\phi_1, \phi_2, 0) - \frac{1}{2}(R(\phi_1, \phi_2, \Delta\lambda) + R(\phi_1, \phi_2, -\Delta\lambda)).$$

If $X(P)$ is longitudinally reversible, that is, $R(\phi_1, \phi_2, \Delta\lambda) = R(\phi_1, \phi_2, -\Delta\lambda)$, we have

$$\gamma(\phi_1, \phi_2, \Delta\lambda) = R(\phi_1, \phi_2, 0) - R(\phi_1, \phi_2, \Delta\lambda).$$

When $\phi_1 = \phi_2$, the above equation reduces to the variogram function on the circle.

3.1.2 Block Circulant Covariance Matrix for Axially Symmetry Processes

Let $X(\phi, \lambda)$ be an axially symmetric process on the sphere. Under the assumption of gridded data X specified above by (3.1), the variance-covariance matrix of $X(\phi, \lambda)$ is a block circulant matrix $C \in \mathsf{R}^{n_l n \;\; n_l n}$ where n_l and n denote the number of latitudes and longitudes, respectively ([Ada17], [Yan13], [Ste07])

$$C = \begin{pmatrix} C_0 & C_1 & C_2 & \cdots & C_{n\ 2} & C_{n\ 1} \\ C_{n\ 1} & C_0 & C_1 & \cdots & C_{n\ 3} & C_{n\ 2} \\ C_{n\ 2} & C_{n\ 1} & C_0 & \cdots & C_{n\ 4} & C_{n\ 3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ C_1 & C_2 & C_3 & \cdots & C_{n\ 1} & C_0 \end{pmatrix}.$$

Here $C_j, j = 0, 1, \cdots, n-1$, is the covariance matrix of $\{X(\phi_i, \lambda_1), i = 1, 2, \cdots, n_l\}$ and $\{X(\phi_i, \lambda_{j+1}), i = 1, 2, \cdots, n_l\}$, that is,

$$C_j = \begin{pmatrix} cov(X(\phi_1, \lambda_1), X(\phi_1, \lambda_{j+1})) & \cdots & cov(X(\phi_1, \lambda_1), X(\phi_{n_l}, \lambda_{j+1})) \\ cov(X(\phi_2, \lambda_1), X(\phi_1, \lambda_{j+1})) & \cdots & cov(X(\phi_2, \lambda_1), X(\phi_{n_l}, \lambda_{j+1})) \\ \vdots & \vdots & \vdots \\ cov(X(\phi_{n_l}, \lambda_1), X(\phi_1, \lambda_{j+1})) & \cdots & cov(X(\phi_{n_l}, \lambda_1), X(\phi_{n_l}, \lambda_{j+1})) \end{pmatrix}.$$

Moreover, for any two latitudes ϕ_P, ϕ_Q, we have

$$\begin{aligned} & cov(X(\phi_P, \lambda_1), X(\phi_Q, \lambda_{i+1})) = R(\phi_P, \phi_Q, i\delta) \\ \neq \; & cov(X(\phi_Q, \lambda_1), X(\phi_P, \lambda_{i+1})) = R(\phi_Q, \phi_P, i\delta), \end{aligned}$$

where $\delta = \frac{2\pi}{n}$. That is, C_i may not be symmetric. However, $C_i = C_n^T{}_i$ because

$$\begin{aligned} & cov(X(\phi_P, \lambda_1), X(\phi_Q, \lambda_{i+1})) = R(\phi_P, \phi_Q, i\delta) \\ = \; & R(\phi_Q, \phi_P, (n-i)\delta) = cov(X(\phi_Q, \lambda_1), X(\phi_P, \lambda_{n\ i+1})). \end{aligned}$$

3.1.3 Symmetric Block Circulant Matrix with Symmetric Blocks for Longitudinally Reversible Processes

We can extend the above work when the underlying process is a longitudinally reversible process. Longitudinally reversible process is a special case of an axially

symmetric process where $R(\phi_P, \phi_Q, \Delta\lambda) = R(\phi_P, \phi_Q, -\Delta\lambda)$. Hence,

$$\begin{aligned} & cov(X(\phi_P, \lambda_1), X(\phi_Q, \lambda_{i+1})) = R(\phi_P, \phi_Q, i\delta) = R(\phi_P, \phi_Q, -i\delta) \\ = \quad & R(\phi_Q, \phi_P, i\delta) = cov(X(\phi_Q, \lambda_1), X(\phi_P, \lambda_{i+1})). \end{aligned}$$

This above derivation shows that all C_i are symmetric blocks with $C_i = C_{n\ \ i}$.

$$\begin{aligned} & cov(X(\phi_P, \lambda_1), X(\phi_Q, \lambda_{i+1})) = R(\phi_P, \phi_Q, i\delta) \\ = \quad & R(\phi_P, \phi_Q, (n-i)\delta) = cov(X(\phi_P, \lambda_1), X(\phi_Q, \lambda_{n\ \ i+1})). \end{aligned}$$

Hence, the block circulant covariance matrix of longitudinally reversible processes is given by

$$C = \begin{pmatrix} C_0 & C_1 & C_2 & \cdots & C_2 & C_1 \\ C_1 & C_0 & C_1 & \cdots & C_3 & C_2 \\ C_2 & C_1 & C_0 & \cdots & C_4 & C_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ C_2 & C_3 & C_4 & \cdots & C_0 & C_1 \\ C_1 & C_2 & C_3 & \cdots & C_1 & C_0 \end{pmatrix},$$

with $C_i = C_i^T$.

3.1.4 Cross Covariance Estimator on the Sphere

According to [Ada17], the MOM (Method of Moments) cross covariance estimator between two points with longitudinal difference of $\Delta\lambda$ for an axially symmetric process

can be written

$$\begin{aligned}
\hat{R}_{12}(\Delta\lambda) &= \hat{R}(\phi_1, \phi_2, \Delta\lambda) \\
&= \frac{1}{n}\sum_{i=1}^{n}(X(\phi_1, (i-1)\delta + \Delta\lambda) - \bar{X}_{\phi_1})(X(\phi_2, (i-1)\delta + \Delta\lambda) - \bar{X}_{\phi_2}) \\
&= \frac{1}{n}\sum_{i=1}^{n} X(\phi_1, (i-1)\delta + \Delta\lambda)X(\phi_2, (i-1)\delta) - \bar{X}_{\phi_1}\bar{X}_{\phi_2} \\
&= \vec{X}^T A(\Delta\lambda)\vec{X}.
\end{aligned}$$

Here

$$\begin{aligned}
\bar{X}_{\phi_i} &= \frac{1}{n}\sum_{j=1}^{n} X(\phi_i, \lambda_j), \quad i = 1, 2, \\
\vec{X} &= (X(\phi_1, \lambda_1), X(\phi_2, \lambda_1), X(\phi_1, \lambda_2), X(\phi_2, \lambda_2), \cdots, X(\phi_1, \lambda_n), X(\phi_2, \lambda_n))^T,
\end{aligned}$$

and $A(\Delta\lambda)$ is a block circulant matrix given by

$$A(\Delta\lambda) = \frac{1}{n}\text{circ}\left(\begin{pmatrix} 0 & \frac{1}{n} \\ 0 & 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 & 1-\frac{1}{n} \\ 0 & 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 & \frac{1}{n} \\ 0 & 0 \end{pmatrix}\right)$$

According to [Ada17], $A(\Delta\lambda)$ can be unitarily diagonalized as following

$$A(\Delta\lambda) = P\text{diag}(S_1^{(A)}, S_2^{(A)}, \cdots, S_n^{(A)})P \ ,$$

where P is a unitary matrix satisfying

$$PP \ = P \ P = I_n$$

with

$$S_j^{(A)} = \sum_{m=0}^{n\ 1} \omega_j^m A_m,$$

where $A_m, m = 0, 1, \cdots, n-1$ represents the 2×2 blocks in $A(\Delta\lambda)$. (See more details in [Ada17]).

3.1.5 Cross Variogram Estimator on the Sphere

Based on [Ada17], the MOM cross variogram estimator for an axially symmetric process can be written

$$\begin{aligned}
\hat{\gamma}_{12} &= \hat{\gamma}(\phi_1, \phi_2, \Delta\lambda) \\
&= \frac{1}{2n} \sum_{i=1}^{n} (X(\phi_1, (i-1)\delta + \Delta\lambda) - (X(\phi_1, (i-1)\delta)) \\
&\quad \times (X(\phi_2, (i-1)\delta + \Delta\lambda) - X(\phi_2, (i-1)\delta)),
\end{aligned}$$

where $A(\Delta\lambda)$ is a block-symmetric circulant matrix. For illustration, we give the detailed expression of $A(\Delta\lambda)$ when $n = 6$.

If $n = 6$, we have the following expressions for $A(\Delta\lambda)$ with $\Delta\lambda = 0, \pi/3, 2\pi/3, \pi$.

$$
\begin{aligned}
A(0) &= 0_{12\ 12} \\
A(\pi/3) &= \frac{1}{12}\mathrm{circ}(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}). \\
A(2\pi/3) &= \frac{1}{12}\mathrm{circ}(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}), \\
A(\pi) &= \frac{1}{12}\mathrm{circ}(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}).
\end{aligned}
$$

Also, $A(\Delta\lambda)$ has a spectral unitary decomposition as follows.

$$A(\Delta\lambda) = P\mathrm{diag}(S_1^{(A)}, S_2^{(A)}, \cdots, S_n^{(A)})P$$

with

$$S_j^{(A)} = (1 - \cos((j-1)\Delta\lambda)) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad j = 1, 2, \cdots, n.$$

Hence,

$$A(\Delta\lambda) = P\mathrm{diag}(1 - \cos((j-1)\Delta\lambda)) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} P \ .$$

Further expressions of $S_j^{(A)}, j = 1, 2, \cdots, n$ and P when $n = 6$ are given in [Ada17].

3.2 Fourier Transform

The Fourier Transform is a tool that breaks a waveform (a function or signal) into an alternate representation, characterized by sine and cosines. The Fourier Series breaks down a periodic function into the sum of sinusoidal functions. It is the Fourier Transform for periodic functions.

The discrete Fourier transform (DFT) converts a finite sequence of equally spaced samples of a function into an equally long sequence of equally spaced samples of the discrete time Fourier transform (DTFT), which is a function of frequency with complex values. The DFT is the most important discrete transform, used to perform Fourier analysis in many practical applications.

The Fourier transform takes various forms:

Continuous Form: The continuous Fourier transform of a signal $x(t) \in \mathbb{C}, t \in (-\infty, \infty)$, is defined as

$$P(w) = \int_{1}^{1} x(t)e^{\ jwt} dt$$

and its inverse given by

$$x(t) = \frac{1}{2\pi} \int_{1}^{1} P(w)e^{jwt} dw.$$

Fourier Series: The Fourier Transform of a function can be derived as a special case of the Fourier Series when the period, $T \to \infty$. Start with the Fourier Series synthesis equation.

$$x(t) = \sum_{n=}^{+1} c_n e^{jn\omega_o t}$$

where c_n is given by the Fourier Series analysis equation,

$$c_n = \frac{1}{T} \int_{1}^{1} x(t) e^{\ jn\omega_0 t} dt$$

which can be rewritten as following.

$$Tc_n = \int_{1}^{1} x(t) e^{\ jn\omega_0 t} dt.$$

As $T \to \infty$ the fundamental frequency, $\omega_0 = 2\pi/T$, becomes extremely small and the quantity $n\omega_0$ becomes a continuous quantity that can take on any value. So we define a new variable $\omega = n\omega_0$; we also let $X(\omega) = Tc_n$. Making these substitutions in the previous equation yields the analysis equation for the Fourier transform.

Discrete Fourier Transform: Finite DFT plays a key role throughout this dissertation work. According to the [Brock91], let $\vec{x} = (x_1, x_2, \cdots, x_n)^T \in \mathsf{C}^n$ be a sequence of n numbers and $\vec{a} = (a_1, a_2, \cdots, a_n)^T$ a vector of real numbers. Then the representation of $\vec{x}$ as a linear combination of harmonics,

$$x_t = \frac{1}{\sqrt{n}} \sum_{0 \ w_j < 2\pi} a_j e^{itw_j}, \qquad t = 1, \cdots, n.$$

Where $w_j = 2\pi j/n$, $j \in \{0, 1, \cdots, (n-1)\} = F_n$ falls in the interval $[0, 2\pi)$. The vectors e_j, $j \in F_n$ are defined by the following

$$\vec{e_j} = n^{\ 1/2}(1, e^{iw_j}, e^{i2w_j}, \cdots, e^{i(n\ \ 1)w_j}).$$

Define the inner product of two vectors $\vec{u} = (u_1, u_2, \cdots, u_n)^T, \vec{v} = (v_1, v_2, \cdots, v_n)^T \in \mathsf{C}^n$ as follows.

$$< \vec{u}, \vec{v} >= \sum_{t=1}^{n} u_t \bar{v}_t.$$

Hence from the properties of the above inner product, one can show that, for $j, k \in F_n$,

$$< \vec{e}_j, \vec{e}_k >= \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

That is $\{\vec{e}_j, j \in F_n\}$ forms a unitary basis of C^n.

The above representation of $\vec{x}$ can be written in vector form as

$$\vec{x} = \sum_{j 2 F_n} a_j \vec{e}_j.$$

We can derive the a_j where $j \in F_n$ by taking the inner product of $\vec{x}$ with $\vec{e}_j$.

$$a_j = \langle \vec{x}, \vec{e}_j \rangle = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t \overline{e^{itw_j}} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{\ itw_j}.$$

In summary, we have

$$a_j = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{\ itw_j} \qquad \text{and} \qquad x_t = \frac{1}{\sqrt{n}} \sum_{j 2 F_n} a_j e^{itw_j}$$

to form a DFT pair.

Now let us consider the following special cases.

- First, when $\vec{a}$ is real, that is, $a_j = \overline{a_j}$ for all $j = 0, 1, 2, \cdots, n-1$. We have

$$\begin{aligned} a_j = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{\ itw_j} = \bar{a}_j = \overline{\frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{\ itw_j}} \\ \Rightarrow \quad \sum_{t=1}^{n} x_t e^{\ itw_j} = \sum_{t=1}^{n} \overline{x_t} e^{itw_j}. \end{aligned}$$

  Expanding the last two summands, and letting $x_t = u_t + iv_t$ and $e^{\ itw_j} = \cos(tw_j) - i\sin(tw_j)$, we have

$$\begin{aligned} 0 &= \sum_{t=1}^{n} ((u_t + iv_t)(\cos(tw_j) - i\sin(tw_j)) - (u_t - iv_t)(\cos(tw_j) + i\sin(tw_j))) \\ &= \sum_{t=1}^{n} 2i(v_t \cos(tw_j) - u_t \sin(tw_j)) \\ \Rightarrow \quad &\sum_{t=1}^{n} (v_t \cos(tw_j) - u_t \sin(tw_j)) = 0. \end{aligned}$$

  Therefore, if we take $x_t = u_t + iv_t$ with both u_t and v_t real, we have

$$\begin{aligned} a_j &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (u_t + iv_t)(\cos(tw_j) - i\sin(tw_j)) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (u_t \cos(tw_j) + v_t \sin(tw_j)) + \sum_{t=1}^{n} (v_t \cos(tw_j) - u_t \sin(tw_j)) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (u_t \cos(tw_j) + v_t \sin(tw_j)). \end{aligned}$$

On the other hand, if we write

$$x_t = u_t + iv_t = \frac{1}{\sqrt{n}} \sum_{j2F_n} a_j e^{itw_j} = \frac{1}{\sqrt{n}} \sum_{j2F_n} a_j(\cos(tw_j) + i\sin(tw_j)),$$

then, $u_t = \frac{1}{n} \sum_{j2F_n} a_j \cos(tw_j)$ and $v_t = \frac{1}{n} \sum_{j2F_n} a_j \sin(tw_j)$. Moreover, we note that $u_t = u_{n\ t}, v_t = -v_{n\ t}$, implying that $x_{n\ t} = \bar{x}_t, t = 1, 2, \cdots, n-1$.

Remark: We will apply the above result when we consider the kriging method for axially symmetric processes in Section 3.3.

- If $x_t, t = 1, 2, \cdots, n$ are real, it is obvious that a_0 is real. We further prove that $a_{n\ j} = \bar{a}_j$ for $j = 1, 2, \cdots, n-1$.

$$a_{n\ j} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{\ itw_{n-j}} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{\ it(2\pi)(n\ \ j)/n} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{it(2\pi)j/n} = \bar{a}_j.$$

Now we consider the following expansion for x_t. For simplicity, we set $n = 2N$.

$$x_t = \frac{a_0}{\sqrt{n}} + \frac{1}{\sqrt{n}} \left(\sum_{j=1}^{N\ 1} a_j e^{itw_j} + \sum_{j=N+1}^{n\ 1} a_j e^{itw_j} \right) + \frac{1}{\sqrt{n}} a_N e^{it\pi} \quad t = 1, \cdots, n.$$

Let $k = n - j$ in the second summation above. We have

$$\sum_{j=N+1}^{n\ 1} a_j e^{itw_j} = \sum_{k=1}^{N\ 1} a_{n\ k} e^{itw_{n-k}} = \sum_{k=1}^{N\ 1} \bar{a}_k e^{\ itw_k}.$$

Hence,

$$\begin{aligned} x_t &= \frac{a_0}{\sqrt{n}} + \frac{1}{\sqrt{n}}\left(\sum_{j=1}^{N}{}^{1} a_j e^{itw_j} + \sum_{j=N+1}^{n}{}^{1} a_j e^{itw_j}\right) + \frac{a_N}{\sqrt{n}} e_{it\pi} \\ &= \frac{a_0}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{j=1}^{N}{}^{1}(a_j e^{itw_j} + \bar{a}_j e^{\ itw_j}) + \frac{1}{\sqrt{n}} a_N (-1)^t \\ &= \frac{a_0}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{j=1}^{N}{}^{1}(a_j e^{itw_j} + \bar{a}_j e^{\ itw_j}) + \frac{1}{\sqrt{n}} a_N (-1)^t. \end{aligned}$$

Note that if n is odd, the last term in the above expression disappears.

If we write $a_j = r_j e^{i\theta_j}$, then we can rewrite the above expression as

$$x_t = \frac{a_0}{\sqrt{n}} + \sqrt{2}\sum_{j=1}^{N}(r_j c_j \cos\theta_j - r_j s_j \sin\theta_j) + \frac{1}{\sqrt{n}} a_N (-1)^t, \qquad t = 1, \cdots, n$$

where for $j = 1, 2, \cdots, N-1$,

$$c_j = \sqrt{\frac{2}{n}}(1, \cos w_j, \cos 2w_j, \cdots, \cos(n-1)w_j)^T,$$

and

$$s_j = \sqrt{\frac{2}{n}}(0, \sin w_j, \sin 2w_j, \cdots, \sin(n-1)w_j)^T.$$

Therefore, now we can establish the orthonormal basis of R^n by $\{e_0, c_1, s_1, \cdots, c_{N_1}, s_{N_1}, e_N\}$, where $e_0 = \frac{1}{\sqrt{n}}(1, 1, \cdots, 1)^T$, and

$e_N = \frac{1}{\sqrt{n}}(-1, 1, -1, 1, \cdots, 1)$ if n is even.

- If both $\vec{a} = (a_0, a_1, \cdots, a_{n\ 1})^T$ and $\vec{x}$ are real, then we have

$$a_j = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t \cos(itw_j) \qquad \text{and} \qquad x_t = \frac{1}{\sqrt{n}} \sum_{j2F_n} a_j \cos(itw_j).$$

The above expression can be simplified further. For simplicity, we assume that $n = 2N$, an even integer. Note that when both $\vec{x}$ and a_j are real, one can deduce that $a_{n\ j} = a_j$ and $x_t = x_{n\ t}$ for all available j and t. Therefore, the above two equations can be simplified as following:

$$\begin{aligned} a_j &= \frac{1}{\sqrt{n}} \left(x_n + 2 \sum_{t=1}^{N\ 1} x_t \cos(tw_j) + x_N(-1)^j \right), \\ x_t &= \frac{1}{\sqrt{n}} \left(a_0 + 2 \sum_{j=1}^{N\ 1} a_j \cos(tw_j) + a_N(-1)^t \right). \end{aligned}$$

If n is odd, the last term in the above expressions disappears.

In the above real value DFT, the orthonormal basis is given by $\{e_0, c_1, \cdots, c_{N\ 1}\}$ and e_N if n is even.

Remark: We will apply the above result when we consider the kriging method for longitudinally reversible processes in Section 3.3.

3.3 Kriging Method Development

In this section, we first provide the general idea for developing kriging methods for axially symmetric processes on the sphere. In particular, we note that an axially symmetric process on the sphere can be decomposed as Fourier series on circles, where the Fourier random coefficients can be expressed as circularly-symmetric possibly complex random processes. Based on the Discrete Fourier transform, ordinary kriging methods have been developed for axially symmetric processes with zero mean, constant mean, and different means on different latitudes, respectively, after which both parametric and non-parametric approaches will be proposed. That will be provided in details in Chapter IV.

3.3.1 Introduction

Let $X(\phi, \lambda)$ be an axially symmetric process on the sphere, where $\phi \in [0, \pi]$ represents the latitude, and $\lambda \in [0, 2\pi]$ represents the longitudes. Based on Proposition 2.5 from [HZR12], an axially symmetric random process with mean zero on the sphere can be written as

$$X(\phi, \lambda) = \sum_{m=-1}^{1} W_m(\phi) e^{im\lambda}, \tag{3.2}$$

where the complex random process $W_m(\phi), \phi \in [0, \pi]$ can be obtained by

$$W_m(\phi) = \frac{1}{2\pi} \int_0^{2\pi} X(\phi, \lambda) e^{-im\lambda} d\lambda,$$

satisfying $E(W_m(\phi_P)\overline{W_n(\phi_Q)}) = \delta_{n,m}C_m(\phi_P, \phi_Q)$. Here $C_m(\phi_P, \phi_Q)$ is possibly a complex covariance function of $W_m(\phi)$. Further, if the Gaussianity of $X(\phi, \lambda)$ is assumed, $\{W_m(\phi), m = 0, \pm 1, \pm 2, \ldots\}$ are independent circularly-symmetric Gaussian complex random process on $\phi \in [0, \pi]$ (Proposition 1 from [VWZ21]).

If the random process $X(\phi, \lambda)$ on the sphere is a longitudinally reversible process, $X(\phi, \lambda)$ can be represented as follows.

$$X(\phi, \lambda) = W_0(\phi) + 2\sum_{m=1}^{1} W_m(\phi)\cos(m\lambda),$$

with

$$W_0(\phi) = \frac{2}{\pi}\int_0^{2\pi} X(\phi, \lambda)d\lambda, \quad W_m(\phi) = \frac{1}{\pi}\int_0^{2\pi} X(\phi, \lambda)\cos(m\lambda)d\lambda, m = 1, 2, \cdots .$$

Here $W_m(\phi), m = 0, 1, \cdots$ are uncorrelated real-valued random processes on $\phi \in [0, \pi]$. If the Gaussianity is further assumed, $W_m(\phi)$ will be independent Gaussian random processes on $\phi \in [0, \pi]$.

3.3.2 Kriging Method Development - Main Idea

Suppose we have finite gridded data values $\{X(\phi_i, \lambda_j), i = 1, 2, \cdots, n_l; j = 1, 2, \cdots, n\}$ where $\lambda_j = (j-1)\delta$ with $\delta = \frac{2\pi}{n}$ but $\phi_i, i = 1, 2, \cdots, n_l$ might not be gridded. The purpose in this section is to provide the kriging method to predict the value $X(\phi_0, \lambda_0)$ where ϕ_0 may not be one of the latitudes $\phi_i, i = 1, 2, \cdots, n_l$.

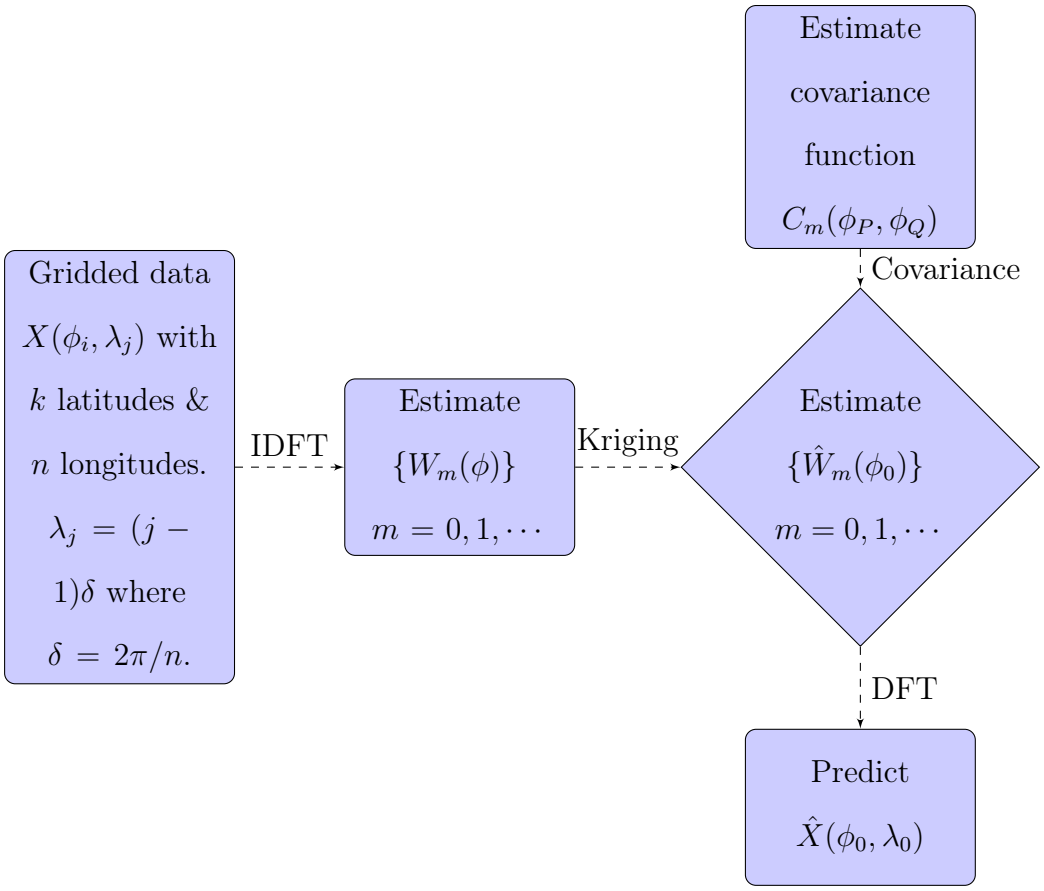

Figure 3. Spatial prediction flow

The above diagram (Figure 3) can be decomposed into the following major steps:

(1) Given gridded data values $\{X(\phi_i, \lambda_j), i = 1, 2, \cdots, n_l; j = 1, 2, \cdots, n\}$, we obtain $W_m(\phi_i), i = 1, 2, \cdots, n_l; m = 0, 1, \cdots$ through an inverse DFT.

(2) Based on the given/estimated gridded data values, we estimate the covariance function $C_m(\cdot, \cdot), m = 0, 1, 2, \cdots$ (details will be discussed in Chapter IV).

(3) Perform kriging to estimate $\{W_m(\phi_0)\}, m = 0, 1, \cdots,$ where ϕ_0 is the given latitude.

(4) Use DFT to obtain the predicted value $\hat{X}(\phi_0, \lambda_0)$ at the new location (ϕ_0, λ_0).

We will illustrate the above kriging flow chart under two separate settings: when $X(\phi, \lambda)$ is longitudinally reversible, and when $X(\phi, \lambda)$ is axially symmetric.

3.3.3 *Kriging under Longitudinal Reversibility*

We assume that $X(\phi, \lambda)$ is a longitudinally reversible process on the sphere. As we note from Section 3.3.1, $W_m(\phi), m = 0, 1, 2, \cdots$ becomes a real-valued uncorrelated random process on $\phi \in [0, \pi]$. We illustrate the above diagram(Figure 3) in more details.

Step 1: Obtaining $\boldsymbol{W_m(\phi), m = 0, 1, 2, \cdots}$, with Inverse DFT

For a fixed latitude ϕ, if we are given finite gridded data values, we have the following DFT expression (take $n = 2N$ for simplicity).

$$X(\phi, \lambda) = \frac{1}{\sqrt{n}} \left(W_0(\phi) + 2 \sum_{m=1}^{N} {}^{1} W_m(\phi) \cos(m\lambda) + \cos(\pi) W_N(\pi) \right), \tag{3.3}$$

and hence based on Section 3.2,

$$W_m(\phi) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X(\phi, \lambda_t) \cos(m\lambda_t), \quad m = 0, 1, \cdots, N.$$

In general, if we are given gridded data on a set of latitudes, denoted as $\{\phi_1, \phi_2, \cdots, \phi_m\}$, we can write the above expression in a matrix form as given below.

$$W = \frac{1}{\sqrt{n}} X e,$$

where X is the gridded data structure given in Section 3.1 and

$$W = \begin{pmatrix} W_0(\phi_1) & W_1(\phi_1) & \cdots & W_{n\ 1}(\phi_1) \\ W_0(\phi_2) & W_1(\phi_2) & \cdots & W_{n\ 1}(\phi_2) \\ W_0(\phi_3) & W_1(\phi_3) & \cdots & W_{n\ 1}(\phi_3) \\ \vdots & \vdots & \vdots & \\ W_0(\phi_m) & W_1(\phi_m) & \cdots & W_{n\ 1}(\phi_m) \end{pmatrix},$$

$$e = \begin{pmatrix} 1 & (\cos\lambda_1) & (\cos 2\lambda_1) & \cdots & 1 \\ 1 & (\cos\lambda_2) & (\cos 2\lambda_2) & \cdots & -1 \\ 1 & (\cos\lambda_3) & (\cos 2\lambda_3) & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (\cos\lambda_n) & (\cos 2\lambda_n) & \cdots & -1 \end{pmatrix}.$$

Step 2: Estimate $C_m(\cdot,\cdot)$ (given in Chapter IV)

Step 3: Kriging Method

Assume that we have the gridded data X given by (3.1). We will predict the data value $X(\phi_0, \lambda_0)$. We have the following two cases.

(1) Latitude $\phi_0 \in \{\phi_1, \phi_2, \cdots, \phi_m\}$. In other words, ϕ_0 is one of the existing latitudes.

This is a straight forward approach. One can easily get the $W_m(\phi)$ from the above equation (3.3) and apply the following equation. More specifically, any data value at latitude ϕ_0 and longitude λ_0 can be written as

$$X(\phi_0, \lambda_0) = \frac{1}{\sqrt{n}}(\hat{W}_0(\phi_0) + 2\sum_{m=0}^{N\ 1} \hat{W}_m(\phi_0)\cos(m\lambda_0) + \hat{W}_N(\phi_0)\cos(N\pi)). \quad (3.4)$$

(2) Latitude $\phi_0 \in [0, \pi]$ but may not be one of the existing latitudes. In this case, we need to predict $W_m(\phi_0), m = 0, 1, 2, \cdots, N$ first (details to be given next), and then we can use the above expression(3.4) to obtain $X(\phi_0, \lambda_0)$.

In principle, ordinary kriging is represented as a linear combination of observed data values [Cre93]. It has the following two assumptions.

(1) Model assumption: The real-valued random process $W(\phi)$ with an unknown constant mean μ can be written as follows.

$$W(\phi) = \mu + \delta(\phi), \quad \phi \in [0, \pi], \mu \in \mathsf{R} \text{ and } \mu \text{ unknown.}$$

where $\delta(\phi)$ is a real-valued process with zero mean but having the same covariance structure.

(2) Predictor assumption: The unobserved $W(\phi_0)$ at an unknown location ϕ_0 is computed by

$$W(\phi_0) = \sum_{i=1}^{n} c_i W(\phi_i), \quad \sum_{i=1}^{n} c_i = 1$$

where $c_i \in \mathsf{R}$ $i = 1, 2, \cdots, n$ are to be determined.

Note

$$E[W(\phi_0)] = E\left(\sum_{i=1}^{n} c_i W(\phi_i)\right) = \sum_{i=1}^{n} c_i E(W(\phi_i)) = \sum_{i=1}^{n} c_i \mu = \mu.$$

From the unbiasedness we can say MSE = Variance.

$$\sigma^2 = E(\mu - W(\phi_0))^2 = E(W(\phi_0) - \sum_{i=1}^{n} c_i W(\phi_i))^2 = Var(W(\phi_0) - \sum_{i=1}^{n} c_i W(\phi_i)).$$

Therefore, for the optimal prediction on the real-valued random field, we need to minimize the mean squared error (MSE), given by $E(W(\phi_0) - \sum_{i=1}^{n} c_i W(\phi_i))^2$. Here, $W(\phi_0)$ and $W(\phi_i)$ can be replaced by the $\delta(\phi_0) + \mu$ and $\delta(\phi_i) + \mu$, respectively, due to the predictor assumption where $\sum_{i=1}^{n} \lambda_i = 1$.

Hence,

$$\begin{aligned}
& E(W(\phi_0) - \sum_{i=1}^{n} c_i W(\phi_i))^2 \\
= \; & E(\delta(\phi_0) - \sum_{i=1}^{n} c_i \delta(\phi_i))^2 \\
= \; & E\delta(\phi_0)^2 - 2E(\delta(\phi_0) \sum_{i=1}^{n} c_i \delta(\phi_i)) + E(\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \delta(\phi_i)\delta(\phi_j)) \\
= \; & E(W(\phi_0) - \mu)^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j E((W(\phi_i) - \mu)(W(\phi_j) - \mu)) \\
& -2 \sum_{i=1}^{n} c_i E((W(\phi_0) - \mu)(W(\phi_i) - \mu)) \\
= \; & C(\phi_0, \phi_0) + \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j C(\phi_i, \phi_j) - 2 \sum_{i=1}^{n} c_i C(\phi_0, \phi_i),
\end{aligned}$$

where $C(\phi_0, \phi_0)$ is the variance value at ϕ_0 and $C(\phi_i, \phi_j)$ is the covariance at two latitudes ϕ_i, ϕ_j. After applying the Lagrange multiplier l subject to $\sum_{i=1}^{n} c_i = 1$, the objective function becomes

$$C(\phi_0, \phi_0) + \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j C(\phi_i, \phi_j) - 2\sum_{i=1}^{n} c_i C(\phi_0, \phi_i) - 2l\left(\sum_{i=1}^{n} c_i - 1\right).$$

Differentiating with respect to $c_1, \cdots, c_n$ and l, and equating to 0, we have

$$\sum_{j=1}^{n} c_j C(\phi_i, \phi_j) - C(\phi_0, \phi_i) - l = 0 \quad i = 1, 2, \cdots, n.$$

and

$$\sum_{i=1}^{n} c_i = 1.$$

We can rewrite the above equations in the following matrix form

$$\begin{bmatrix} A_{n\ \ n} & -\vec{1}_{n\ \ 1} \\ \vec{1}'_{1\ \ n} & 0 \end{bmatrix} \begin{bmatrix} \vec{c}_{n\ \ 1} \\ \mathrm{l} \end{bmatrix} = \begin{bmatrix} \vec{b}_{n\ \ 1} \\ 1 \end{bmatrix}.$$

where $\vec{1}_{n\ \ 1} = (1, 1, \cdots, 1)^T$, $\vec{c}_{n\ \ 1} = (c_1, c_2, \cdots, c_n)^T$, and

$$A = \begin{pmatrix} C(\phi_1, \phi_1) & C(\phi_1, \phi_2) & C(\phi_1, \phi_3) & \cdots & C(\phi_1, \phi_n) \\ C(\phi_2, \phi_1) & C(\phi_2, \phi_2) & C(\phi_2, \phi_3) & \cdots & C(\phi_2, \phi_n) \\ C(\phi_3, \phi_1) & C(\phi_3, \phi_2) & C(\phi_3, \phi_3) & \cdots & C(\phi_3, \phi_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C(\phi_n, \phi_1) & C(\phi_n, \phi_2) & C(\phi_n, \phi_3) & \cdots & C(\phi_n, \phi_n) \end{pmatrix},$$

$$\vec{b} = (C(\phi_1, \phi_0), C(\phi_2, \phi_0), \cdots, C(\phi_n, \phi_0))^T,$$

Optimal Solution

Now we are looking for the solution to the following equation:

$$A_{n\ \ n}\vec{c}_{n\ \ 1} - l\vec{1}_{n\ \ 1} = \vec{b}_{n\ \ 1},$$

subject to $\sum_{i=1}^{n} c_i = 1$.

Hence,

$$\vec{c}_{n\ \ 1} = A^{\ \ 1}[\vec{b}_{n\ \ 1} + l\vec{1}_{n\ \ 1}].$$

But, we know that $\sum_{i=1}^{n} c_i = 1$. Therefore, multiplying the transpose of $\vec{1}^T_{n\ \ 1} = \vec{1}_{1\ \ n}$ by the left side of the above equation gives.

$$1 = \vec{1}_{1\ \ n}A^{\ \ 1}[\vec{b}_{n\ \ 1} + l\vec{1}_{n\ \ 1}], \quad \text{and} \quad l = \frac{[1 - \vec{1}_{1\ \ n}A^{\ \ 1}\vec{b}_{n\ \ 1}]}{\vec{1}_{1\ \ n}A^{\ \ 1}\vec{1}_{n\ \ 1}}.$$

Once we plug the solution for l into the equation $c_{n\ \ 1} = A^{\ \ 1}[\vec{b}_{n\ \ 1} + l\vec{1}_{n\ \ 1}]$, we have,

$$c_{n\ \ 1} = A^{\ \ 1}\left[\vec{b}_{n\ \ 1} + \vec{1}_{n\ \ 1}\frac{[1 - \vec{1}_{1\ \ n}A^{\ \ 1}b]}{\vec{1}_{1\ \ n}A^{\ \ 1}\vec{1}_{n\ \ 1}}\right].$$

For a longitudinally reversible process $X(\phi, \lambda)$, $W_m(\phi)\ \ m = 0, 1, 2, \cdots, n$ is a real process. Apply the above kriging methods to each of $W_m(\phi)$ to obtain the estimates

$\hat{W}_m(\phi_0)$.

$$\hat{W}_m(\phi_0) = \sum_{j=1}^{n} c_j^{(m)} W_m(\phi_j) \quad m = 0, 1, 2, \cdots$$

where $c^{(m)} = A_m{}^1 \left[\vec{b}_m + \vec{1}_n \;\; 1 \frac{[1 \;\; \vec{1}_{1\times n} A_m^{-1} \vec{b}_m]}{\vec{1}_{1\times n} A_m^{-1} \vec{1}_{n\times 1}} \right]$,

$$A_m = \begin{pmatrix} C_m(\phi_1,\phi_1) & C_m(\phi_1,\phi_2) & C_m(\phi_1,\phi_3) & \cdots & C_m(\phi_1,\phi_n) \\ C_m(\phi_2,\phi_1) & C_m(\phi_2,\phi_2) & C_m(\phi_2,\phi_3) & \cdots & C_m(\phi_2,\phi_n) \\ C_m(\phi_3,\phi_1) & C_m(\phi_3,\phi_2) & C_m(\phi_3,\phi_3) & \cdots & C_m(\phi_3,\phi_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_m(\phi_n,\phi_1) & C_m(\phi_n,\phi_2) & C_m(\phi_n,\phi_3) & \cdots & C_m(\phi_n,\phi_n) \end{pmatrix},$$

$$b_m = (C_m(\phi_1,\phi_0), C_m(\phi_2,\phi_0), \cdots, C_m(\phi_n,\phi_0))^T$$

Step 4: Data Prediction

We use the inverse DFT to predict the data $X(\phi_0, \lambda_0)$ from $\hat{W}_m(\phi_0)$.

$$X(\phi_0,\lambda_0) = \frac{1}{\sqrt{n}}(\hat{W}_0(\phi_0) + 2\sum_{m=0}^{N}{}^{1} \hat{W}_m(\phi_0)\cos(m\lambda_0) + \hat{W}_N(\phi_0)\cos(N\pi)).$$

3.3.4 Kriging under Axial Symmetry

We assume that $X(\phi, \lambda)$ is an axially symmetric process on the sphere. As we note from Section 3.3.1, $W_m(\phi), m = 0, 1, 2, \cdots$ becomes a complex-valued uncorrelated random process on $\phi \in [0, \pi]$. We illustrate the above diagram in more details.

Step 1: Obtaining $\boldsymbol{W_m(\phi), m = 0, 1, 2, \cdots}$, with Inverse DFT

Unlike longitudinally reversible process, an axially symmetric random process on the sphere $W_m(\phi), \phi \in [0, \pi]$ is possibly a complex-valued random process. Let $X(P)$ be an axially symmetric random process on the sphere. For each fixed latitude ϕ, let $\vec{X} = (X(\phi, \lambda_1), X(\phi, \lambda_2), \cdots, X(\phi, \lambda_n)$ with $\lambda_1 = 0, \lambda_2 = \delta, \cdots, \lambda_i = (i-1)\delta$ with $\delta = \frac{2\pi}{n}$ be observed gridded data on the sphere. Then from [HZR12], we have

$$X(\phi, \lambda_t) = \frac{1}{\sqrt{n}} \sum_{m=0}^{n\ 1} W_m(\phi) e^{im\lambda_t}, \quad t = 1, 2, \cdots, n.$$

Therefore, using the inverse discrete Fourier transform, we have (for example, [Brock91])

$$W_m(\phi) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X(\phi, \lambda_t) e^{\ im\lambda_t}, \qquad m = 0, 1, 2, \cdots, n-1.$$

For the general case of gridded data with a dimension of $n_l \times n$ (that is, n_l latitudes and n longitudes), the above equation can be rewritten as

$$W = \frac{1}{\sqrt{n}} Xe$$

where

$$X = \begin{pmatrix} X(\phi_1,\lambda_1) & X(\phi_1,\lambda_2) & X(\phi_1,\lambda_3) & \cdots & X(\phi_1,\lambda_n) \\ X(\phi_2,\lambda_1) & X(\phi_2,\lambda_2) & X(\phi_2,\lambda_3) & \cdots & X(\phi_2,\lambda_n) \\ X(\phi_3,\lambda_1) & X(\phi_3,\lambda_2) & X(\phi_3,\lambda_3) & \cdots & X(\phi_3,\lambda_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X(\phi_m,\lambda_1) & X(\phi_m,\lambda_2) & X(\phi_m,\lambda_3) & \cdots & X(\phi_m,\lambda_n) \end{pmatrix},$$

$$W = \begin{pmatrix} W_0(\phi_1) & W_1(\phi_1) & \cdots & W_{n\ 1}(\phi_1) \\ W_0(\phi_2) & W_1(\phi_2) & \cdots & W_{n\ 1}(\phi_2) \\ W_0(\phi_3) & W_1(\phi_3) & \cdots & W_{n\ 1}(\phi_3) \\ \vdots & \vdots & \vdots & \\ W_0(\phi_m) & W_1(\phi_m) & \cdots & W_{n\ 1}(\phi_m) \end{pmatrix},$$

$$e = \begin{pmatrix} e^{\ i1\lambda_1} & e^{\ i2\lambda_1} & e^{\ i3\lambda_1} & \cdots & e^{\ in\lambda_1} \\ e^{\ i1\lambda_2} & e^{\ i2\lambda_2} & e^{\ i3\lambda_2} & \cdots & e^{\ in\lambda_2} \\ e^{\ i1\lambda_3} & e^{\ i2\lambda_3} & e^{\ i3\lambda_3} & \cdots & e^{\ in\lambda_3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e^{\ i1\lambda_n} & e^{\ i2\lambda_n} & e^{\ i3\lambda_n} & \cdots & e^{\ in\lambda_n} \end{pmatrix}.$$

Therefore, for any latitude ϕ_p we can obtain $W_m(\phi_p), m = 1, 2, \cdots, n$ from $X(\phi_p, \lambda_i), i = 1, 2, \cdots, n$.

Step 2: Estimate $C_m(\cdot,\cdot)$ (given in Chapter IV)

In the rest of this section, we will present the kriging method for predicting random values on unobserved locations when the underlying process is axially symmetric.

Step 3: Kriging for the complex $\boldsymbol{W_m(\phi)}$

Making inferences on unobserved values of the complex random process $W(\cdot)$ from the complex data $W(\phi) = (W(\phi_1), \cdots, W(\phi_n))^T$ observed at known spatial locations $\{\phi_1, \cdots, \phi_n\}$ is called kriging on a complex-valued random field. Similar to the case under the longitudinally reversible process, we have two different cases of predictions.

(1) Latitude $\phi_0 \in \{\phi_1, \phi_2, \cdots, \phi_m\}$. In other words, ϕ_0 is an existing latitude. This is a straight forward prediction. One can easily get the $W_m(\phi)$ from the data based on the inverse DFT and then apply the following equation. More specifically, any data value $X(\phi_0, \lambda_0)$ to be predicted at the unobserved location (ϕ_0, λ_0) can be written as

$$\hat{X}(\phi_0, \lambda_0) = W_0(\phi_0) + 2\sum_{m=1}^{m\ \ 1}[W_m^{Re}(\phi_0)\cos(m\lambda_0) - W_m^{Im}(\phi_0)\sin(m\lambda_0)]$$

where $W_m(\phi_0) = W_m^{Re}(\phi_0) + iW_m^{Im}(\phi_0)$ is a possibly complex-valued random process on $\phi \in [0, \pi]$.

(2) Latitude $\phi_0 \in [0, \pi]$ but not one of the existing latitudes. We consider this case in more detail.

Spatial predictions on an axially symmetric process has the following two assumptions.

(1) Model assumption: The complex-valued random process $W(\phi)$ with an unknown constant mean μ can be written as the following.

$$W(\phi) = \mu + \delta(\phi), \quad \phi \in [0, \pi], \mu \in \mathsf{R} \text{ and } \mu \text{ unknown}$$

where $\delta(\phi)$ is a complex-valued process with zero mean but has the same covariance structure.

(2) Predictor assumption: The unobserved $W(\phi_0)$ at an unknown location ϕ_0 is computed by

$$W(\phi_0) = \sum_{i=1}^{n} b_i W(\phi_i), \quad \sum_{i=1}^{n} b_i = 1$$

where $b_i \in \mathsf{C}$ $i = 1, 2, \cdots, n$ are to be determined.

According to the [Sand15], we denote the complex-valued random field $W(\phi)$ at unknown location ϕ as the following

$$W(\phi) = X(\phi) + \mathit{i}\mathsf{Y}(\phi),$$

where $X(\phi)$ and $Y(\phi)$ represent the real and imaginary components of $W(\phi)$, respectively.

Therefore, the complex-valued random variable $W(\phi_0)$ at an unknown location ϕ_0 can be computed

$$\begin{aligned} W(\phi_0) &= \sum_{i=1}^{n} b_i W(\phi_i) \\ &= \sum_{j=1}^{n} (b_j^{Re} + ib_j^{Im})(X(\phi_j) + iY(\phi_j)) \\ &= \sum_{j=1}^{n} [b_j^{Re} X(\phi_j) - b_j^{Im} Y(\phi_j)] + i \sum_{j=1}^{n} [b_j^{Re} Y(\phi_j) + b_j^{Im} X(\phi_j)] \\ &= X^T b^{Re} - Y^T b^{Im} + i(X^T b^{Im} + Y^T b^{Re}) \end{aligned}$$

where $\vec{X} = (X(\phi_1), X(\phi_2), \cdots, X(\phi_n))^T$, $\vec{Y} = (Y(\phi_1), Y(\phi_2), \cdots, Y(\phi_n))^T$, $\vec{b}^{Re} = (b_1^{Re}, b_2^{Re}, \cdots, b_n^{Re})^T, \vec{b}^{Im} = (b_1^{Im}, b_2^{Im}, \cdots, b_n^{Im})^T$ are the real and imaginary components of the predictors, respectively.

However, we have

$$E[W(\phi_0)] = E(\sum_{i=1}^{n} b_i W(\phi_i)) = \sum_{i=1}^{n} b_i E(W(\phi_i)) = \sum_{i=1}^{n} b_i \mu = \mu.$$

Hence $W(\phi_0)$ is unbiased only if $\sum_{i=1}^{n} b_i = 1$, which implies that

$$\sum_{i=1}^{n} b_i^{Re} = 1, \quad \sum_{i=1}^{n} b_i^{Im} = 0.$$

Note that

$$\text{MSE} = \text{Variance} + \text{Bias}^2,$$

and from the unbiasedness we have MSE = Variance, giving

$$\sigma^2 = E|\mu - W(\phi_0)|^2 = E|W(\phi_0) - \sum_{i=1}^{n} b_i W(\phi_i)|^2 = Var(W(\phi_0) - \sum_{i=1}^{n} b_i W(\phi_i)).$$

Therefore, for the optimal prediction on the complex-valued random field, we need to minimize

$$E \mid W(\phi_0) - \sum_{i=1}^{n} b_i W(\phi_i) \mid^2 = E(W(\phi_0) - \sum_{i=1}^{n} b_i W(\phi_i))\overline{(W(\phi_0) - \sum_{i=1}^{n} b_i W(\phi_i))}.$$

Here, $W(\phi_0)$ and $W(\phi_i)$ can be replaced by $\delta(\phi_0) + \mu$ and $\delta(\phi_i) + \mu$ respectively, due to the predictor assumption that $\sum_{i=1}^{n} b_i = 1$. Hence,

$$\begin{aligned}
& E(\mid W(\phi_0) - \sum_{i=1}^{n} b_i W(\phi_i) \mid^2) \\
= \; & E(\mid \delta(\phi_0) - \sum_{i=1}^{n} b_i \delta(\phi_i) \mid^2) \\
= \; & E \mid \delta(\phi_0) \mid^2 - 2E(\delta(\phi_0)\overline{\sum_{i=1}^{n} b_i \delta(\phi_i)}) + E(\sum_{i=1}^{n}\sum_{j=1}^{n} b_i \overline{b_j} \delta(\phi_i)\overline{\delta(\phi_j)}) \\
= \; & E \mid W(\phi_0) - \mu \mid^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} b_i \overline{b_j} E(W(\phi_i) - \mu)\overline{(W(\phi_j) - \mu)} \\
& -2\sum_{i=1}^{n} b_i E(W(\phi_0) - \mu)\overline{(W(\phi_i) - \mu)} \\
= \; & C(\phi_0, \phi_0) + \sum_{i=1}^{n}\sum_{j=1}^{n} b_i \overline{b_j} C(\phi_i, \phi_j) - 2\sum_{i=1}^{n} b_i C(\phi_0, \phi_i).
\end{aligned}$$

Here $C(\phi_i, \phi_j)$ is the complex covariance function of W.

Finally, minimizing the expression using derivatives with respect to $b_1, b_2, \cdots, b_n$ after

adding the complex Langrange multiplier M, we have

$$\sum_{j=1}^{n} b_j C(\phi_i, \phi_j) + M = C(\phi_0, \phi_i) \quad i = 1, 2, \cdots, n$$

and

$$\sum_{i=1}^{n} b_i^{Re} = 1, \quad \sum_{i=1}^{n} b_i^{Im} = 0.$$

The optimal equation can be rewritten as

$$\sum_{j=1}^{n} (b_i^{Re} + i b_i^{Im})(C^{Re}(\phi_i, \phi_j) + i C^{Im}(\phi_i, \phi_j)) + (M^{Re} + i M^{Im}) = C^{Re}(\phi_0, \phi_i) + i C^{Im}(\phi_0, \phi_i),$$

where $C^{Re}, C^{Im}, M^{Re}, M^{Im}$ are the real and complex components of the corresponding C, M respectively. Therefore, we can split the above n equations into $2n$ equations.

$$C^{Re}(\phi_i, \phi_j) b^{Re} - C^{Im}(\phi_i, \phi_j) b^{Im} + M^{Re} = C^{Re}(\phi_0, \phi_i)$$
$$C^{Im}(\phi_i, \phi_j) b^{Re} + C^{Re}(\phi_i, \phi_j) b^{Im} + M^{Im} = C^{Im}(\phi_0, \phi_i)$$

Further, we obtain this condition from the unbiasedness.

$$\sum_{i=1}^{n} b_i^{Re} = 1, \quad \sum_{i=1}^{n} b_i^{Im} = 0$$

Altogether we have $2n + 2$ unknowns and $2n + 2$ equations.

Matrix Setup

The following matrix setup designed to get the solution for the coefficients($\vec{b}$) and complex Langrange multiplier.

$$\begin{bmatrix} C^{Re} & -C^{Im} & \vec{1}_{n\ 1} & \vec{0}_{n\ 1} \\ C^{Im} & C^{Re} & \vec{0}_{n\ 1} & \vec{1}_{n\ 1} \\ \vec{1}_{1\ n} & \vec{0}_{1\ n} & 0 & 0 \\ \vec{0}_{1\ n} & \vec{1}_{1\ n} & 0 & 0 \end{bmatrix} \begin{bmatrix} \vec{b}_{n}^{Re}{}_{1} \\ \vec{b}_{n}^{Im}{}_{1} \\ M^{Re} \\ M^{Im} \end{bmatrix} = \begin{bmatrix} \vec{C}^{Re}(\phi_0, \phi_i)_{n\ 1} \\ \vec{C}^{Im}(\phi_0, \phi_i)_{n\ 1} \\ 1 \\ 0 \end{bmatrix}$$

where C^{Re}, C^{Im} are the real and imaginary components of the covaraince matrix, respectively.

Let us take

$$A = \begin{bmatrix} C^{Re} & -C^{Im} \\ C^{Im} & C^{Re} \end{bmatrix}_{2n\ \ 2n}, \quad B = \begin{bmatrix} \vec{1}_{n\ 1} & \vec{0}_{n\ 1} \\ \vec{0}_{n\ 1} & \vec{1}_{n\ 1} \end{bmatrix}, \quad C = \begin{bmatrix} \vec{1}_{1\ n} & \vec{0}_{1\ n} \\ \vec{0}_{1\ n} & \vec{1}_{1\ n} \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

But, A and $D - CA^{\ 1}B$ are non-singular, hence from the analytic inversion formula we have the $(2n+2) \times (2n+2)$ matrix given by

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{1} = \begin{bmatrix} A^{\ 1} + A^{\ 1}B(D - CA^{\ 1}B)^{\ 1}CA^{\ 1} & -A^{\ 1}B(D - CA^{\ 1}B)^{\ 1} \\ -(D - CA^{\ 1}B)^{\ 1}CA^{\ 1} & (D - CA^{\ 1}B)^{\ 1} \end{bmatrix}.$$

Furthermore, the matrix A is symmetric and C^{Re} is non-singular. Finally, the solution matrix is

$$\begin{bmatrix} \vec{b}_n^{Re}{}_1 \\ \vec{b}_n^{Im}{}_1 \\ M^{Re} \\ M^{Im} \end{bmatrix}$$

$$= \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} \vec{C}^{Re}(\phi_0, \phi_i)_n{}_1 \\ \vec{C}^{Im}(\phi_0, \phi_i)_n{}_1 \\ 1 \\ 0 \end{bmatrix}.$$

We can then solve the above equation to obtain $\vec{b} = \vec{b}^{Re} + i\vec{b}^{Im}$. Hence, the predicted value $\hat{W}(\phi_0)$ is given by

$$\hat{W}(\phi_0) = \sum_{j=1}^{n} (b_j^{Re} + ib_j^{Im}) W(\phi_j).$$

For the prediction on the axially symmetric process, we repeat the above procedure for each $m = 0, 1, 2, \cdots, n-1$ to have the following

$$\hat{W}_m(\phi_0) = \sum_{j=1}^{n} (b_{mj}^{Re} + ib_{mj}^{Im}) W_m(\phi_j) \quad m = 0, 1, 2, \cdots, n-1.$$

Step 4: Obtain the Predicted Value $\boldsymbol{X(\phi_0, \lambda_0)}$

Finally, the data value at the unobserved location (ϕ_0, λ_0) can be obtained by

$$\hat{X}(\phi_0, \lambda_0) = \frac{1}{\sqrt{n}} \sum_{m=0}^{n} {}^{1}[\hat{W}_m^{Re}(\phi_o)\cos(m\lambda_0) - \hat{W}_m^{Im}(\phi_0)\sin(m\lambda_0)].$$

3.3.5 Justification for Kriging (IPE)

[Yan13] discussed the justification for kriging performance. Suppose that $\hat{X}_0(\phi, \lambda)$ is the predicted value at location s using the true covariance function C_0, and $\hat{X}_i(\phi, \lambda)$ is the predicted value with covariance function C_i. Let $e_0(\phi, \lambda) = X(\phi, \lambda) - \hat{X}_0(\phi, \lambda)$ and $e_i(\phi, \lambda) = X(\phi, \lambda) - \hat{X}_i(\phi, \lambda)$ be the prediction error, respectively.

Ee_0^2 is the mean squared prediction error (MSPE) of the best linear unbiased predictor or the kriging variance, which is computed using the true covariance function. Ee_i^2 is the actual variance of the prediction error. Then the Increase in Prediction Error (IPE) is given by

$$\frac{Ee_i^2(s)}{Ee_0^2(s)} = \frac{E(X(\phi, \lambda) - \hat{X}_i(\phi, \lambda))^2}{E(X(\phi, \lambda) - \hat{X}_0(\phi, \lambda))^2}.$$

Remark: A smaller IPE value indicates a better kriging performance for the corresponding approach.

CHAPTER IV

COVARIANCE MODEL ESTIMATION

In this chapter, our main goal is to estimate the covariance function that is critical in the kriging development. More specifically, we propose both parametric and non-parametric approaches in estimating the true covariance function. We perform simulations to evaluate the accuracy of our approaches for axially symmetric processes on the sphere.

In the parametric approach, we specify a model with some parameters that are to be determined. These parameters are normally estimated/adjusted from the data. The advantage of using parametric approaches is that only a few unknown parameters are to be estimated, which could be computationally efficient if a correct model is specified. However, in practice mostly we lack of knowledge about the underlying process of the model. This leads to model misspecification. For instance, if the underlying true model is exponential but we incorrectly assume that it is a spherical model then we will end up with making incorrect statistical inferences and conclusions.

Unlike a parametric approach, the non-parametric approach does not make an assumption for the underlying covariance model. It has a potential advantage to accurately fit a wide range of possible models. However, non-parametric approaches often suffer from the major disadvantage that a large number of observations, compared with the number of observations needed for parametric approaches, might be

needed in order to obtain an accurate model estimate.

In this chapter, we will introduce both parametric and non-parametric approaches when estimating the covariance models. In Section 4.1, we discuss one parametric approach: least squared error estimation method (LSE). The performances of the LSE method is compared through simulation studies in Section 4.2. In Section 4.3 we apply the Wavelet-Galerkin approximation to obtain the covariance model. Simulation studies are also conducted to investigate the performance of proposed methods in Section 4.4. Some discussions are given in Section ??.

4.1 Parametric Approaches

4.1.1 Least Squared Method

The least squared method is a classical parametric approach, where we select the parameters based on the least sum of squared differences between the true and estimated semivariogram values, as the Method of Moments (MOM) semivariogram estimator is unbiased (for example, [Van16]). More specifically, let $R(\phi_P, \phi_Q, \Delta\lambda; \vec{\theta}) \equiv R(\vec{\theta})$ be the underlying covariance model with parameters $\vec{\theta}$ to be determined. Hence the semivariogram function can be given below.

$$\gamma(\vec{\theta}) \equiv \gamma(\phi_P, \phi_Q, \Delta\lambda; \vec{\theta}) = R(\phi_P, \phi_Q, 0; \vec{\theta}) - \frac{1}{2}\left(R(\phi_P, \phi_Q, \Delta\lambda; \vec{\theta}) + R(\phi_Q, \phi_P, \Delta\lambda; \vec{\theta})\right).$$

With the given gridded data, we calculate the estimated semivariogram values through

the following formula.

$$\begin{aligned}\hat{\gamma}(\phi_P, \phi_Q, \Delta\lambda) &= \frac{1}{2n}\sum_{i=1}^{n}\left(X(\phi_P, (i-1)\delta + \Delta\lambda) - X(\phi_P, (i-1)\delta)\right) \\ &\quad \times \left(X(\phi_Q, (i-1)\delta + \Delta\lambda) - X(\phi_Q, (i-1)\delta)\right).\end{aligned}$$

We choose the parameters $\hat{\vec{\theta}}$ that minimize the sum of squared differences between true and estimated semivariogram values, that is,

$$\hat{\vec{\theta}} = \text{argmin}_{\vec{\theta}} \sum_{\text{all pairs } (\phi_i, \phi_j)} \sum_{\Delta\lambda} (\gamma(\phi_i, \phi_j, \Delta\lambda; \vec{\theta}) - \hat{\gamma}(\phi_i, \phi_j, \Delta\lambda))^2.$$

We then use the covariance model with optimal parameter values to perform kriging. Here is the algorithm for this method.

Algorithm 1 (Pseudo-code)

- Step-1: Generate data based on the given covariance function $R(P, Q; \vec{\theta})$ with a specified parameter vector $\vec{\theta}$.

- Step-2: Select optimal parameters $\hat{\vec{\theta}}$ of the parametric covariance function based on the LSE method.

  (1) Assume we are given data $\{X(\phi_i, \lambda_j), 1 \leq i \leq n_l, 1 \leq j \leq n\}$ on n_l latitudes and n gridded longitudes.

  (2) For specified parameter values, compute the theoretical variogram estimator $\gamma(\phi_1, \phi_2, \Delta\lambda; \vec{\theta})$ for all pairs of ϕ_i, ϕ_j & and all longitude differences $\Delta\lambda$

$$\begin{aligned}\gamma(\phi_1, \phi_2, \Delta\lambda; \vec{\theta}) &= R(\phi_1, \phi_2, 0; \vec{\theta}) \\ &\quad -\frac{1}{2}(R(\phi_1, \phi_2, \Delta\lambda; \vec{\theta}) + R(\phi_1, \phi_2, -\Delta\lambda; \vec{\theta})).\end{aligned}$$

  (3) Compute the estimated variogram estimator for all pairs of ϕ_1, ϕ_2 and all longitude differences $\Delta\lambda$.

  (4) Select the optimal parameter values $\hat{\vec{\theta}}$ such that

$$\hat{\vec{\theta}} = \text{argmin}_{\vec{\theta}} \sum_{\text{all pairs } (\phi_i, \phi_j)} \sum_{\Delta\lambda} (\gamma(\phi_i, \phi_j, \Delta\lambda; \vec{\theta}) - \hat{\gamma}(\phi_i, \phi_j, \Delta\lambda))^2.$$

- Step-3: Use the optimal parameters of the given covariance function to perform kriging.

4.2 Parametric Approaches - Simulation Studies

In this section, we conduct simulations to compare the performance of the proposed parametric approaches. We will compare both the accuracy of parameter estimates and the IPE values.

4.2.1 Covariance Function Selection

In this research, we use the following covariance function that was proposed by [VWZ21]:

$$R(\phi_P, \phi_Q, \Delta\lambda) = \tilde{C}(\phi_p, \phi_q)\frac{1-p^2}{1-2p\cos\alpha+p^2},$$

$$\tilde{C}(\phi_p, \phi_q) = C_1(C_2 - e^{\ aj\phi_P j} - e^{\ aj\phi_Q j} + e^{\ aj\phi_P \ \ \phi_Q j})$$

where $\alpha = \Delta\lambda + u(\phi_p - \phi_q) \in [0, 2\pi], C_1 \geq 0, C_2 \geq 1, a \geq 0, u \in \mathsf{R}, p \in (0,1)$.

There are five parameters C_1, C_2, a, p and u in the model, where C_1, C_2, a, p are scaling parameters and u is a parameter where $u = 0$ indicates the longitudinal reversibility of a random process. For illustration purposes, we will set $\vec{\theta} = (a, C_2)^T$ as parameters to be determined.

According to the Proposition 2 in [HZR12], we have

$$C_m(\phi_p, \phi_q) = \frac{1}{2\pi}\int_{\pi}^{\pi} R(\phi_P, \phi_Q, \Delta\lambda)e^{\ im\Delta\lambda}d\Delta\lambda = \tilde{C}(\phi_P, \phi_Q)p^m e^{imb} \tag{4.1}$$

with $b = u(\phi_P - \phi_Q)$. Note that $C_0(\phi_p, \phi_q) = \tilde{C}(\phi_p, \phi_q)$.

4.2.2 Data Generation

To generate the gridded data values $\{X(\phi_i, \lambda_j), 1 \leq i \leq n_l, 1 \leq j \leq n\}$ with n_l latitudes and n equal-spacing longitudes, we use the classical data generation method. Note that the covariance matrix for an axially symmetric process is block circulant, and as a special case, it becomes a symmetric block circulant matrix with symmetric blocks for a longitudinally reversal process. Let $R(\phi_P, \phi_Q, \Delta\lambda; \vec{\theta})$ be the covariance matrix of an axially symmetric process $X(\phi, \lambda)$, and let $\vec{Z}$ represent an *i.i.d.* standard normal random vector. Then the gridded data, denoted as a matrix X as given by (3.1), can be obtained by

$$X = R^{\frac{1}{2}}(\phi_P, \phi_Q, \Delta\lambda; \vec{\theta}) \times \vec{Z}.$$

Here $R^{1/2}$ is the square root matrix of R, which can be easily obtained through Singular Value Decomposition (SVD). Finally, we consider both axially symmetric and longitudinally reversible processes with the following three mean trend scenarios: zero means, global constant means, and different constant means on different latitudes.

In our simulation, we set the following initial parameters that were also considered in [VWZ21] for data generation.

Table 1. Parametric approach data generation parameter settings

Parameter	C_1	C_2	a	u	p
initial value	1	2	1	1	0.5

For all cases considered in this dissertation, we generate a grid of 11×10 data structure X on latitudes $(\pi/11, 2\pi/11, \cdots, \pi)$ and longitudes $(0, 2\pi/10, 4\pi/10, \cdots, 18\pi/10)$. We consider the axially symmetric process by setting $u = 1$ and the longitudinally reversible process with $u = 0$. Furthermore, for each of three mean cases, we use the following mean structures.

(1) Zero mean: $E(X(\phi, \lambda)) = 0$.

(2) Constant mean: $E(X(\phi, \lambda)) = 100$.

(3) Different means on different latitudes: We set mean values incrementally with latitudes with values $(110, 120, \cdots)$.

Finally we repeat the above process iteratively with iteration number $= 100$.

4.2.3 Results and Conclusions - Parametric Approaches

The following tables describe the bias and mean squared error values when estimating parameters $\vec{\theta} = (a, C_2)^T$ for LSE approach under the axially symmetric process for all the three mean cases.

Table 2. The bias and MSE values for estimating $\vec{\theta} = (a, C_2)^T$ for LSE approach when the underlying process is axially symmetric (iteration = 100)

$n = 10, n_l = 11$		True	LSE Estimates (Bias(MSE))
Zero mean	a	1	0.14(0.78)
	C_2	2	0.49 (1.09)
Constant mean	a	1	0.14(0.78)
	C_2	2	0.49 (1.09)
Different means	a	1	0.14(0.78)
	C_2	2	0.49 (1.09)

Table 3. The bias and MSE values for estimating $\vec{\theta} = (a, C_2)^T$ for LSE approach when the underlying process is longitudinally reversible (iteration = 100)

$n = 10, n_l = 11$		True	LSE Estimates (Bias(MSE))
Zero mean	a	1	0.23(1.72)
	C_2	2	0.02 (0.59)
Constant mean	a	1	0.24(1.72)
	C_2	2	0.04 (0.58)
Different means	a	1	0.24(1.73)
	C_2	2	0.04 (0.58)

According to the above tables, it should be noted that despite the fact that parameter C_2 in the LSE method when the underlying process is longitudinally reversible is close to the true value, the estimate for the parameter a has a considerable difference with its true parameter value. Meanwhile, parameter a in the LSE method under axially symmetric process is close to the true value. Moreover, one can note that the results for all three cases seem the same. This might be due to the reason that theoretically the same constant on each latitude is subtracted in the estimation of empirical variograms.

One point that can be taken away from the tables is that even with the change of mean structure from the case of zero mean to the case of different means on different latitudes, the parameter value estimates, in particular if they are obtained from the LSE approach, are stable in terms of bias and MSE values. It might be interesting to see if this observation holds in the future if the mean structure gets more complicated, even with non-constant mean trends.

Now we evaluate the performance of our parametric approach when it is used for prediction. Here we use the Increase in Prediction Error (IPE) value (see [Yan13]) to justify the kriging performance. The idea is to calculate a ratio that takes the average overall prediction error based on the estimated model related to that based on the true model. Here we compare the IPE values based on three mean structures. The following two tables provide the IPE values of the proposed kriging methods based on the LSE method when the underlying process is either axially symmetric or longitudinally reversible with three mean structures; that is, zero means, constant

means, and different means on different latitudes.

Table 4. IPE values for LSE approach when the underlying process is axially symmetric with three mean structures.

Axial Symmetry	LSE (Mean(MSE))
Zero mean	58.32 (13366)
Constant mean	58.32 (13366)
Different means	125 (11665)

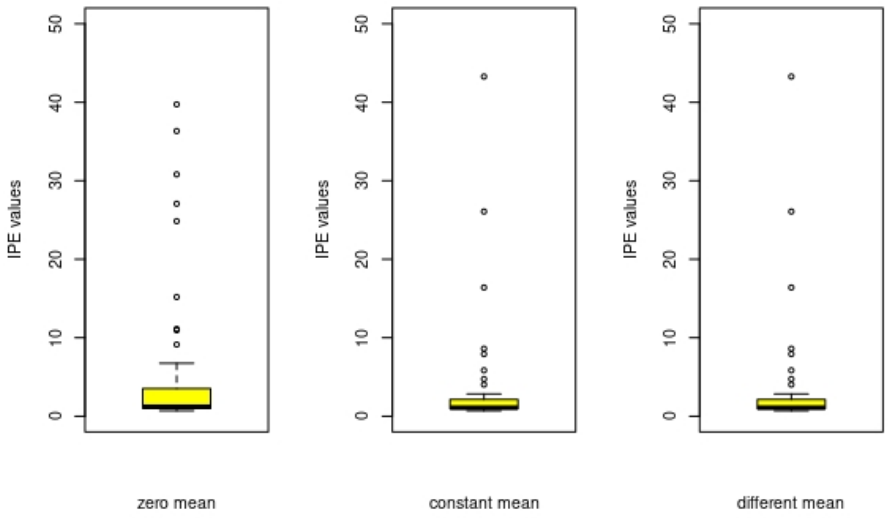

Figure 4. Boxplots for IPE values with axially symmetric processes under zero means (left plot), constant means (center plot), and different means (right plot).

Table 5. IPE values when the underlying process is longitudinally reversible with three mean structures.

Longitudinally Reversible	LSE (Mean(MSE))
Zero mean	27.09(9.66)
Constant mean	5.03(9.66)
Different means	1.44(9.66)

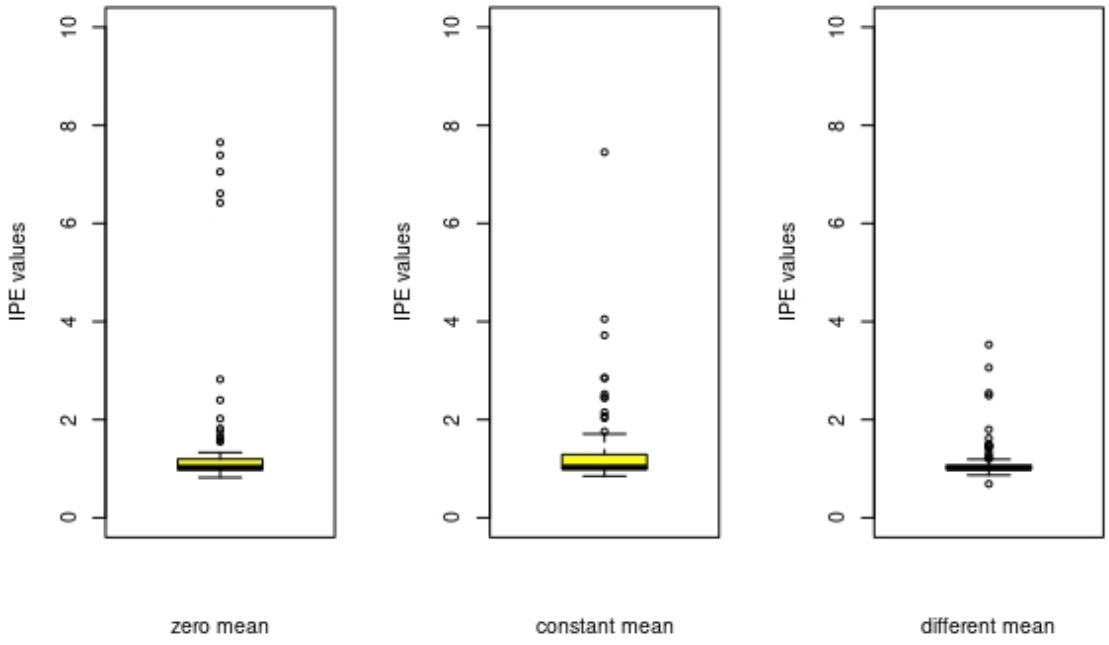

Figure 5. Boxplots for IPE values with longitudinally reversible process under zero means (left plot), constant means (center plot), and different means (right plot).

Based on the above tables and box-plots, we can conclude that the LSE method seems performing fine for all scenarios considered. It should be noted that among those 100 simulations some of the IPE values are unusually large, especially when the underlying process is axially symmetric. These few unusual values heavily impact the calculation of the mean and MSE. To see these more closely, we add box-plots of IPE values under all three cases to illustrate the distribution of IPE values.

From these two set of box-plots (Figures 4 & 5), it is obvious to see that we have some IPE values which are too far from the central box of the distribution. In addition, it seems that LSE methods perform better for longitudinally reversible processes on each of the mean structures than the counterpart for axially symmetric processes. Even though the parameter estimation seems good, the performance of kriging in terms of mean and MSE of IPE values are not justifiable. One potential issue is when the underlying process is axially symmetric the covariance matrix is possibly complex. To overcome this issue, one can try a different justification method for kriging performance (other than IPE) or try a different parametric estimation method.

Finally we investigate the robustness of our parametric approaches in predicting values with increasing data variation. We assume that the underlying process is Gaussian. We then construct a 95% confidence theoretical band to check whether the predicted values fall within the band. For illustration, we consider LSE method under axially symmetric (AS) and longitudinally reversible (AR) processes with a constant mean of 100, where the data are generated with standard deviation values of 1, 10, and 20. The following three sets of plots give the confidence band, the true value,

and the estimated value vs the longitudinal index that is on the predicted latitude.

From the below graphs, the predicted data values are very close to the true values, and they are all well within the theoretical bands for all three cases considered. Therefore, our approaches seem to be robust with increasing variance.

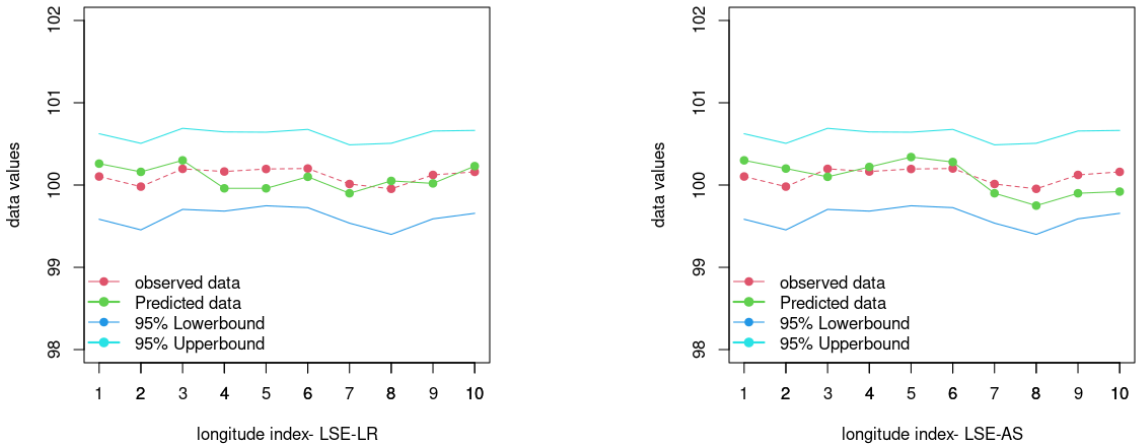

Figure 6. 95% confidence band, estimated and true values vs the longitudinal index for both axially symmetric (AS) and longitudinally reversible (LR) processes with constant mean of 100 and standard deviation of 1.

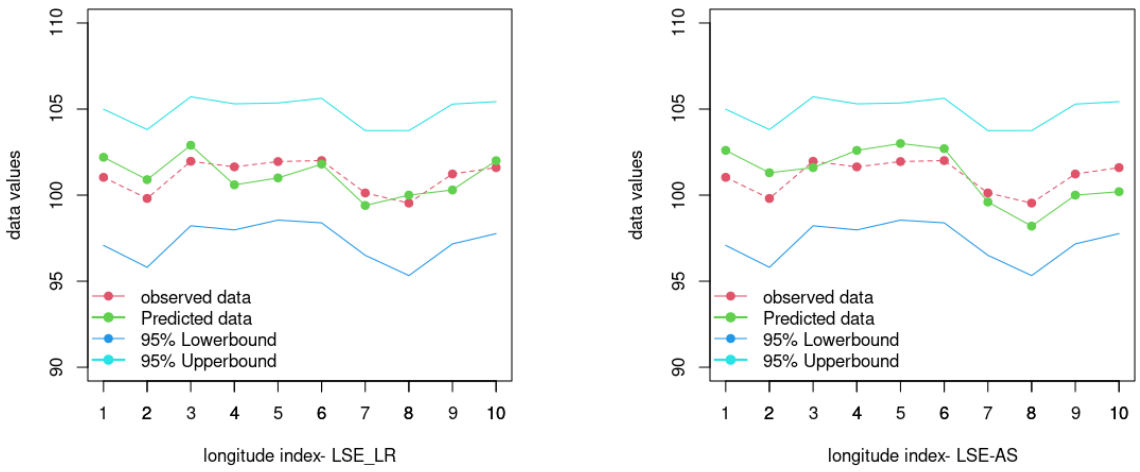

Figure 7. 95% confidence band, estimated and true values vs the longitudinal index for both axially symmetric (AS) and longitudinally reversible (LR) processes with constant mean of 100 and standard deviation of 10.

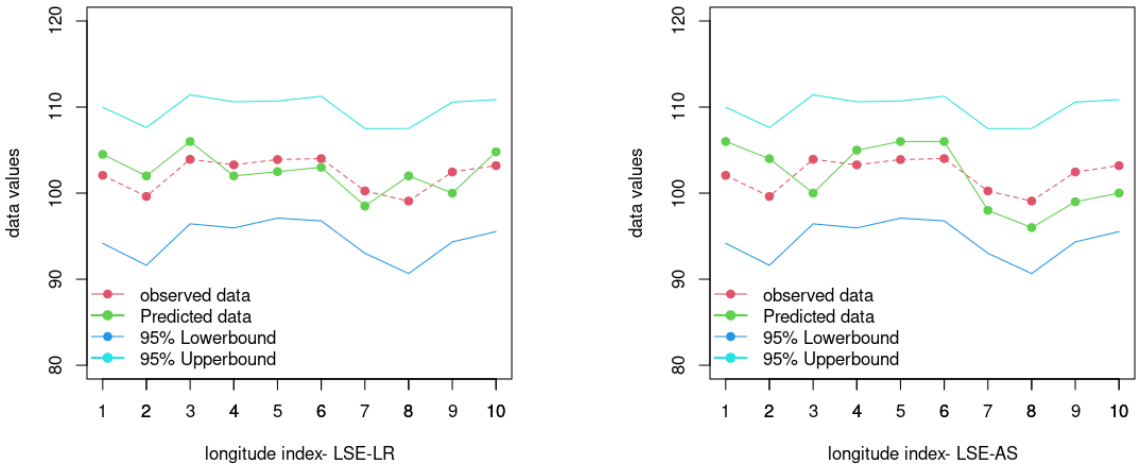

Figure 8. 95% confidence band, estimated and true values vs the longitudinal index for both axially symmetric (AS) and longitudinally reversible (LR) processes with constant mean of 100 and standard deviation of 20.

4.3 Non-parametric Approaches

In this section, we discuss how the two non-parametric approaches are developed and their performance for the axially symmetric process and its special case longitudinally reversible process with three mean structures considered in Section 4.2.

4.3.1 Method Development

In this subsection, we explore an unbiased estimator of the covariance function, which is critical for our non-parametric approaches. According to [Van16], the Method of Moments (MOM) estimator for the cross covariance $R(\phi_P, \phi_Q, \Delta\lambda)$ is given by

$$\hat{R}_M(\phi_P, \phi_Q, \Delta\lambda) = \frac{1}{n}\sum_{i=1}^{n}(X(\phi_P, \lambda_i + \Delta\lambda) - \bar{X_P})(X(\phi_Q, \lambda_i) - \bar{X_Q}) \tag{4.2}$$

where $\bar{X}_P, \bar{X}_Q$ are the latitude means and $\Delta\lambda \in \left\{0, \frac{2\pi}{n}, \frac{4\pi}{n}, \cdots, \frac{(n\ \ 1)2\pi}{n}\right\}$.

If we assume that $E(X(\phi, \lambda)) = \mu_\phi$, that is, the mean of $X(\phi, \lambda)$ is a constant on each latitude ϕ, it has been shown (for example, Proposition 4.2 in [Van16]) that,

$$E(\hat{R}_M(\phi_P, \phi_Q, \Delta\lambda)) = R(\phi_P, \phi_Q, \Delta\lambda) - C_0(\phi_P, \phi_Q). \tag{4.3}$$

That is, the MOM cross covariance estimator $\hat{R}_M(\phi_P, \phi_Q, \Delta\lambda)$ is biased with the constant shift given by $C_0(\phi_P, \phi_Q) = cov(\bar{X}_P, \bar{X}_Q)$. However, it seems impossible to estimate $C_0(\phi_P, \phi_Q) = cov(\bar{X}_P, \bar{X}_Q)$ based on only one copy of the gridded data structure $\mathcal{C}$. Hence, we also make the following assumption.

Assumption: We assume that we have t ***i.i.d.*** copies of gridded data structures, denoted as $\mathcal{D}_k = \{X_k(\phi_i, \lambda_j) : i = 1, 2, \cdots, n_l, j = 1, 2, \cdots, n\}, k = 1, 2, \cdots, t$.

Remark: In the Microwave Sounding Units (MSU) data, average monthly temperature starting in 1978. If we are interested in the spatial process for a particular month, say August, we may assume that we have the roughly independent copies of August data since 1978. Of course, as a part of future research, one should assume those copies have certain temporal dependence, which form a spatio-temporal process.

Now let us introduce the following terminologies. For each latitude $i, i = 1, 2, \cdots, n_l$ on the k^{th} gridded data, we have

$$\bar{X}_{ki} = \frac{1}{n} \sum_{j=1}^{n} X_k(\phi_i, \lambda_j)$$

as the average on latitude i on the k^{th} copy, and

$$\bar{\bar{X}}_i = \frac{1}{t} \sum_{k=1}^{t} \bar{X}_{ki}$$

is the overall average on latitude i. Based on the t ***i.i.d.*** copies $\{\mathcal{D}_k, k = 1, 2, \cdots, t\}$, we propose the following estimator of $C_0(\phi_P, \phi_Q)$ on two latitudes ϕ_P and ϕ_Q, respectively,

$$\hat{C}_0(\phi_P, \phi_Q) = \frac{1}{t-1} \sum_{k=1}^{t} \left(\bar{X}_{kP} - \bar{\bar{X}}_P \right) \left(\bar{X}_{kQ} - \bar{\bar{X}}_Q \right). \tag{4.4}$$

The properties of $\hat{C}_0(\phi_P, \phi_Q)$ are given in the following proposition.

Proposition 4.1: $\hat{C}_0(\phi_P, \phi_Q)$ is an unbiased estimator of $C_0(\phi_P, \phi_Q)$. Moreover, $\hat{C}_0(\cdot, \cdot)$ is non-negative definite.

Proof: We first note that, letting $E(X(\phi_P, \lambda)) = \mu_P$ be the constant mean of $X(\phi_P, \lambda)$ on latitude ϕ_P,

$$\begin{aligned} E(\hat{C}_0(\phi_P, \phi_Q)) &= E\left[\frac{1}{t-1}\sum_{k=1}^{t}\left(\bar{X}_{kP} - \bar{\bar{X}}_P\right)\left(\bar{X}_{kQ} - \bar{\bar{X}}_Q\right)\right] \\ &= E\left[\frac{1}{t-1}\sum_{k=1}^{t}\left((\bar{X}_{kP} - \mu_P) + (\mu_P - \frac{\sum_{k=1}^{n}\bar{X}_{kP}}{n})\right)\right. \\ &\quad \left.\times\left((\bar{X}_Q - \mu_Q) + (\mu_Q - \frac{\sum_{k=1}^{t}\bar{X}_{kQ}}{t})\right)\right] \\ &= I + II + III + IV. \end{aligned}$$

Here we have the following results.

$$\begin{aligned} I &= \frac{1}{t-1}\sum_{k=1}^{t}E(\bar{X}_{kP} - \mu_P)(\bar{X}_{kQ} - \mu_Q) = \frac{t}{t-1}cov(\bar{X}_P, \bar{X}_Q). \\ II &= \frac{1}{t-1}\sum_{k=1}^{t}E\left((\bar{X}_{kP} - \mu_P)(\mu_Q - \bar{\bar{X}}_Q)\right) \\ &= \frac{t}{t-1}E\left(\frac{1}{t}\sum_{k=1}^{t}(\bar{X}_{kP} - \mu_P)(\mu_Q - \bar{\bar{X}}_Q)\right) \\ &= \frac{t}{t-1}E\left((\bar{\bar{X}}_P - \mu_P)(\mu_Q - \bar{\bar{X}}_Q)\right) = -\frac{t}{t-1}cov(\bar{\bar{X}}_P, \bar{\bar{X}}_Q). \end{aligned}$$

$$
\begin{aligned}
III &= \frac{1}{t-1}\sum_{i=1}^{t} E\left((\mu_P - \bar{\bar{X}}_X)((\bar{X}_{kQ} - \mu_Q)\right) \\
&= \frac{t}{t-1} E\left((\mu_P - \bar{\bar{X}}_P)(\bar{\bar{X}}_Q - \mu_Q)\right) = -\frac{t}{t-1} cov(\bar{\bar{X}}_P, \bar{\bar{X}}_Q). \\
IV &= \frac{1}{t-1}\sum_{k=1}^{t} E\left((\mu_P - \bar{\bar{X}}_P)(\mu_Q - \bar{\bar{X}}_Q)\right) = \frac{t}{t-1} cov(\bar{\bar{X}}_P, \bar{\bar{X}}_Q).
\end{aligned}
$$

Note that

$$
\begin{aligned}
cov(\bar{\bar{X}}_P, \bar{\bar{X}}_Q) &= cov\left(\frac{1}{t}\sum_{k=1}^{t}\bar{X}_{kP}, \frac{1}{t}\sum_{s=1}^{t}\bar{X}_{sQ}\right) \\
&= \frac{1}{t^2}\sum_{k=1}^{t}\sum_{s=1}^{t} cov\left(\bar{X}_{kP}, \bar{X}_{sQ}\right) \\
&= \frac{1}{t^2}\sum_{k=1}^{t} cov\left(\bar{X}_{kP}, \bar{X}_{kQ}\right) = \frac{1}{t} cov\left(\bar{X}_P, \bar{X}_Q\right).
\end{aligned}
$$

The last equality is obtained with generic mean processes on latitudes ϕ_P and ϕ_Q, respectively. Hence,

$$
\begin{aligned}
E(\hat{C}_0(\phi_P, \phi_Q)) &= I + II + III + IV \\
&= \frac{t}{t-1} cov(\bar{X}_P, \bar{X}_Q) - \frac{1}{t-1} cov\left(\bar{X}_P, \bar{X}_Q\right) \\
&\quad -\frac{1}{t-1} cov\left(\bar{X}_P, \bar{X}_Q\right) + \frac{1}{t-1} cov\left(\bar{X}_P, \bar{X}_Q\right) \\
&= cov(\bar{X}_P, \bar{X}_Q) = C_0(\phi_P, \phi_Q).
\end{aligned}
$$

That is, $\hat{C}_0(\phi_P, \phi_Q)$ is an unbiased estimator of $C_0(\phi_P, \phi_Q)$. Next we want to prove $\hat{C}_0(\cdot, \cdot)$ is non-negative definite. That is, for any real constants a_i and latitudes $\phi_i, i =$

$1, 2, \cdots, M$,

$$\begin{aligned}
& \sum_{i=1}^{M}\sum_{j=1}^{M} a_i a_j \hat{C}_0(\phi_i, \phi_j) \\
= & \sum_{i=1}^{M}\sum_{j=1}^{M} a_i a_j \frac{1}{t-1}\sum_{k=1}^{t}\left(\bar{X}_{ki} - \bar{\bar{X}}_i\right)\left(\bar{X}_{kj} - \bar{\bar{X}}_j\right) \\
= & \frac{1}{t-1}\sum_{k=1}^{t}\sum_{i=1}^{M}\sum_{j=1}^{M} a_i a_j \left(\bar{X}_{ki} - \bar{\bar{X}}_i\right)\left(\bar{X}_{kj} - \bar{\bar{X}}_j\right) \\
= & \frac{1}{t-1}\sum_{k=1}^{t}\left(\sum_{i=1}^{M} a_i \left(\bar{X}_{ki} - \bar{\bar{X}}_i\right)\right)\left(\sum_{j=1}^{M} a_j \left(\bar{X}_{kj} - \bar{\bar{X}}_j\right)\right) \\
= & \frac{1}{t-1}\sum_{k=1}^{t}\left(\sum_{i=1}^{M} a_i (\bar{X}_{ki} - \bar{\bar{X}}_i)\right)^2 \geq 0.
\end{aligned}$$

Hence, $\hat{C}_0(\cdot, \cdot)$ is non-negative definite.

Now we propose the following estimator of $R(\phi_P, \phi_Q, \Delta\lambda)$.

$$\hat{R}(\phi_P, \phi_Q, \Delta\lambda) = \hat{R}_M(\phi_P, \phi_Q, \Delta\lambda) + \hat{C}_0(\phi_P, \phi_Q).$$

We then have the following properties of $\hat{R}(\phi_P, \phi_Q, \Delta\lambda)$.

Proposition 4.2: $\hat{R}(\phi_P, \phi_Q, \Delta\lambda)$ is an unbiased estimator of $R(\phi_P, \phi_Q, \Delta\lambda)$. It is also non-negative definite.

Proof: The unbiasedness of $\hat{R}(\phi_P, \phi_Q, \Delta\lambda)$ can be obtained from the unbiasedness of $\hat{C}_0(\phi_P, \phi_Q)$ of $C_0(\phi_P, \phi_Q)$ and (4.3). To prove the non-negative positiveness of $\hat{R}(\phi_P, \phi_Q, \Delta\lambda)$, it is sufficient to prove the non-negative definiteness of $\hat{R}_M(\phi_P, \phi_Q, \Delta\lambda)$, which can also be obtained along the same lines as those for proving the non-negative

definiteness of $\hat{C}_0(\phi_P, \phi_Q)$.

Remark: For a zero mean case, the estimator $\hat{R}_M(\phi_P, \phi_Q, \Delta\lambda) = \frac{1}{n}\sum_{i=1}^{n} X(\phi_P, \lambda_i + \Delta\lambda)X(\phi_Q, \lambda_i)$ is the unbiased estimator of $R(\phi_P, \phi_Q, \Delta\lambda)$.

4.3.2 The Wavelet-Galerkin Method

In this dissertation, we apply the Wavelet-Galerkin method to approximate the covariance function, which is critical to our kriging approach. Here we provide an outline of Wavelet-Galerkin method that was developed by [Arachchige21].

Approximating real-valued positive definite covariance functions has become a central part of a geospatial statistics because an accurate approximation of the covariance (kernel) function is crucial for parametric inferences and optimal spatial prediction. Furthermore, investigating structures and properties of complex-valued covariance functions has received special attention in many research fields, for instance, in engineering, complex-valued kernels for complex-valued signals, and in spatial statistics, complex-valued covariance functions for axially symmetric random processes on the sphere. However, there are only minimal studies can be found for approximating complex-covariance functions.

Different types of wavelets and their properties have been widely studied in recent years. In this work we start with a particular wavelet, namely, Haar wavelets. Haar wavelets are known as the simplest basis form of the Daubechies family. On the other

hand, the Galerkin method is considered as a type of expansion method, as eigenfunctions can be written as a linear combination of finite orthogonal basis functions. A linkage of these two concepts: Wavelet-Galerkin method for approximating eigenfunctions and eigenvalues first appeared in [PHQ02] paper. The Galerkin method is a type of an expansion method where any function can be written as a linear combination of finite basis functions where we solve the Fredlohom homogeneous equation.

According to [Arachchige21], Wavelet basis functions are capable of capturing the local information of the approximating functions specially when the eigenfunction is inhomogeneous. Moreover, [Arachchige21] proposed an approach using orthogonal basis functions to approximate a possibly complex-valued covariance function. The ultimate goal of using the Wavelet-Galerkin method in kriging is approximating the covariance function that provides dependency of random processes among observed and unobserved locations. In this work, we used the proposed Wavelet-Galerkin method by [Arachchige21] for our non-parametric kriging approaches, where we approximate the true covariance function $C_m(\phi_P, \phi_Q)$. More explicitly, we use the Wavelet-Galerkin method for two purposes. First. we use the Wavelet-Galerkin method to approximate $R(\phi_P, \phi_Q, \Delta\lambda)$ from $\hat{R}(\phi_P, \phi_Q, \Delta\lambda)$ which is obtained in Section 4.3.4.In addition, we can obtain the estimated $\hat{C}_m(\phi_P, \phi_Q)$ from $\hat{R}(\phi_P, \phi_Q, \Delta\lambda)$ through inverse DFT, and then apply the Wavelet-Galerkin method to approximate the true $C_m(\phi_P, \phi_Q)$.

4.3.3 Outline of the Method

According to Mercer's theorem, a bounded continuous complex-valued continuous, and Hermitian covariance function can be written as follows

$$C(t,s) = \sum_{i=1}^{\infty} \eta_i \; f_i(t) \; \overline{f_i(s)}, \quad a \leq t, s \leq b, \tag{4.5}$$

where $\eta_i \geq 0, i = 1, 2, 3, \ldots$ are the eigenvalues satisfying $\sum_{i=1}^{\infty} \eta_i < \infty$ and $f_i(\cdot)$ are corresponding complex-valued orthonormal and square intergratable eigenfunctions of the covariance function $C(\cdot, \cdot)$. Here $\overline{f_i(s)}$ denotes the complex conjugate of $f_i(s)$. In general, the leading terms in Mercer's theorem capture the main features of the covariance functions, so a truncated expansion of (4.5) can be expressed as $C_N(t,s) = \sum_{i=1}^{N} \eta_i f_i(t)\overline{f_i(s)}$. It can be proved that $C_N(t,s)$, $0 \leq t, s \leq b$ is Hermitian and positive definite ([Arachchige21]). Analytical solutions for the eigenvalues and eigenfunctions can be found by solving the following Fredlohom homogeneous integral equation of a second kind

$$\int_a^b C_N(t,s) f_k(t) dt = \eta_k f_k(t).$$

According to [Arachchige21], the function $f_k(t)$ can be expressed as following.

$$f_k(t) = \sum_{i=0}^{N} \overset{1}{d_i^k} \psi_i(t) = \vec{\Psi}^T(t) \vec{D_k}, t \in \mathsf{R}, d_i^k \in \mathsf{C}$$

$k \in 1, \ldots, N-1$ where, $\vec{D_k}^T = (d_1^k, d_2^k, \ldots, d_N^k) \in \mathsf{C}^N$ is the set of complex-valued wavelet coefficients and $\vec{\Psi}^T(t) = (\psi_1(t), \psi_2(t), \ldots, \psi_N(t)) \in \mathsf{R}^N$ are real-valued Haar

wavelets. Furthermore, $C_N(t,s)$ can be represented in matrix form as follows.

$$C_N(t,s) = F(t)^T \Lambda F(s) = \Psi(\vec{t})^T D \ \Lambda D \Psi(\vec{s}), \tag{4.6}$$

where $F(t)^T = (f_1(t), f_2(t), \ldots, f_N(t))$, for $f_k(t) = \Psi(\vec{t})^T H^{1/2} \vec{\hat{D}}_k$ and Λ is a diagonal matrix with N number of eigenvalue entries and D is the conjugate transpose of D. All of the quantities in (4.6) can be obtained through solving a system of linear equations. For details, see [Arachchige21].

4.3.4 Non-parametric Approaches

Recall from [HZR12], the covariance function $R(\phi_P, \phi_Q, \Delta\lambda)$ can be represented as follows

$$R(\phi_P, \phi_Q, \Delta\lambda) = \sum_{m=-1}^{1} C_m(\phi_P, \phi_Q) e^{im\Delta\lambda},$$

where

$$C_m(\phi_P, \phi_Q) = \frac{1}{2\pi} \int_{-\pi}^{\pi} R(\phi_P, \phi_Q, \Delta\lambda) e^{-im\Delta\lambda} d\Delta\lambda.$$

The above Fourier transform relationship between $R(\phi_P, \phi_Q, \Delta\lambda)$ and $C_m(\phi_P, \phi_Q)$ becomes the basis for our nonparametric approaches in this section.

Given gridded data $\{X(\phi_i, \lambda_j), i = 1, 2, \cdots, n_l, j = 1, 2, \cdots, n\}$ on the sphere, we further assume that we have t *i.i.d.* copies of gridded data structures $\{X_k(\phi_i, \lambda_j), i = 1, 2, \cdots, n_l, j = 1, 2, \cdots, n, k = 1, 2, \cdots, t\}$. From Section 4.3.1, we have the following

unbiased estimator $\hat{R}(\phi_P, \phi_Q, \Delta\lambda)$ of $R(\phi_P, \phi_Q, \Delta\lambda)$ given below.

$$\hat{R}(\phi_P, \phi_Q, \Delta\lambda) = \hat{R}_M(\phi_P, \phi_Q, \Delta\lambda) + \hat{C}_0(\phi_P, \phi_Q),$$

where $\hat{R}_M(\phi_P, \phi_Q, \Delta\lambda)$ and $\hat{C}_0(\phi_P, \phi_Q)$ are given by (4.2) and (4.4), respectively.

Non-parametric approach 1:

In the first nonparametric approach, we first obtain $\hat{C}_m(\phi_p, \phi_q)$ through the inverse DFT

$$\hat{C}_m(\phi_p, \phi_q) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{R}(\phi_p, \phi_q, \Delta\lambda_i) e^{\ im\Delta\lambda_i} \quad m = 0, 1, \cdots .$$

Once we obtain $\hat{C}_m(\phi_p, \phi_q)$, we can use the Wavelet-Galerkin method to approximate the function $C_m(\phi_p, \phi_q)$ to be used for kriging.

Algorithm 2 (Pseudo-code)

- **Step-1**: Given gridded data structure $\{X(\phi_i, \lambda_j), i = 1, 2, \cdots, n_l, j = 1, 2, \cdots, n\}$.
- **Step-2**: Obtain covariance function $C_m(\phi_P, \phi_Q)$ $m = 0, 1, 2, \cdots$.
  (1) Compute the MOM estimator $\hat{R}(\phi_p, \phi_q, \Delta\lambda)$ and $\hat{cov}(\bar{X}_P, \bar{X}_Q)$ from the data for all pairs of ϕ_P, ϕ_Q and all longitude difference $\Delta\lambda$.
  (2) Perform DFT to obtain $\hat{C}_m(\phi_P, \phi_Q)$ for all pairs of ϕ_P, ϕ_Q and $m = 1, 2, \cdots$.
  (3) Apply Wavelet-Galerkin method to derive the $C_m(\phi_P, \phi_Q)$ for all pairs of ϕ_P, ϕ_Q and $0, m = 0, 1, 2, \cdots$.
- **Step-3**: Use $\{C_m(\phi_P, \phi_Q)\}_{m=0,1,2,}$ to perform kriging.

Non-parametric approach 2

Based on the unbiased estimator $\hat{R}(\phi_P, \phi_Q, \Delta\lambda)$ of $R(\phi_P, \phi_Q, \Delta\lambda)$, we perform the Wavelet-Galerkin method to estimate $R(\phi_P, \phi_Q, \Delta\lambda)$.

$$\hat{R}(\phi_p, \phi_q, \Delta\lambda) \xrightarrow{\text{wavelet-Garlekin}} R_{WG}(\phi_p, \phi_q, \Delta\lambda).$$

Therefore, through the inverse Fourier transform, $\hat{C}_m(\phi_p, \phi_q)$ can be obtained through

$$C_m(\phi_p, \phi_q) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{WG}(\phi_p, \phi_q, \Delta\lambda_t) e^{\ im\Delta\lambda_i} \quad m = 0, 1, \cdots .$$

Once we obtain $C_m(\phi_p, \phi_q)$ we can then apply the covariance function on kriging.

Algorithm 3 (Pseudo-code)

- **Step-1** : Given gridded data structure $X(\phi_i, \lambda_j), i = 1, 2, \cdots, n_l, j = 1, 2, \cdots, n$.
- **Step-2** : Obtain covariance function $C_m(\phi_P, \phi_Q)$ $m = 0, 1, 2, \cdots$.
  (1) Compute the MOM estimator $\hat{R}(\phi_p, \phi_q, \Delta\lambda)$ from the data for all pairs of ϕ_P, ϕ_Q and all longitude difference $\Delta\lambda$.
  (2) Perform Wavelet-Galerkin method to obtain $R_{WG}(\phi_p, \phi_q, \Delta\lambda)$

$$\hat{R}(\phi_p, \phi_q, \Delta\lambda) \xrightarrow{\text{wavelet-Galerkin}} R_{WG}(\phi_p, \phi_q, \Delta\lambda).$$

  (3) Apply DFT and estimate the $C_m(\phi_p, \phi_q)$ function from the $R_{WG}(\phi_p, \phi_q, \Delta\lambda)$.
- **Step-3** : Use $\{C_m(\phi_P, \phi_Q)\}_{m=0,1,2,}$ to perform kriging.

4.4 Nonparametric Approaches - Simulation Studies

In the non-parametric approach, we generate data with the same procedure as that for the parametric approach discussed in Section 4.2.2. In addition, we consider both axially symmetric process and longitudinally reversible process with three different mean structures: zero mean, constant mean, and different means on different latitudes, on which we investigate the performance of both non-parametric approaches. We use the covariance function $R(P, Q)$ that was proposed by [Van16] for the data generation (Section 4.2.2). We set the initial parameters below.

Table 6. Non-parametric approach data generation parameter settings

Parameter	C_1	C_2	a	u	p
initial value	1	1	1	1	0.5

The parameter u is set to be 0 if the process is longitudinally reversible and any non-zero for an axially symmetric process. For the convenience of the Wavelet-Galerkin algorithm, we generate the grid data with 16 latitudes and 16 longitudes in our simulation as the Wavelet-Galerkin approximation is very easy to work with when the number of latitudes and longitudes are 2^n.

Remark: If either the number of latitudes or longitudes is not equivalent to 2^n then we can't apply our method directly. Algorithm might need to be generalized.

We first investigate how the estimated $\hat{C}_m(\phi_P, \phi_Q)$ from both non-parametric approaches approximates the true $C_m(\phi_P, \phi_Q)$ for each $m = 0, 1, \cdots,$. We compare the accuracy of approximation through computing the bias between the estimated and the true covariance function for various m values under the two non-parametric approaches. The comparisons are considered over two pairs of of latitudes (10 & 150 and 80 & 100), which are used to explore how the accuracy of approximation changes with the closeness of two latitudes. The results can be seen from the following two plots.

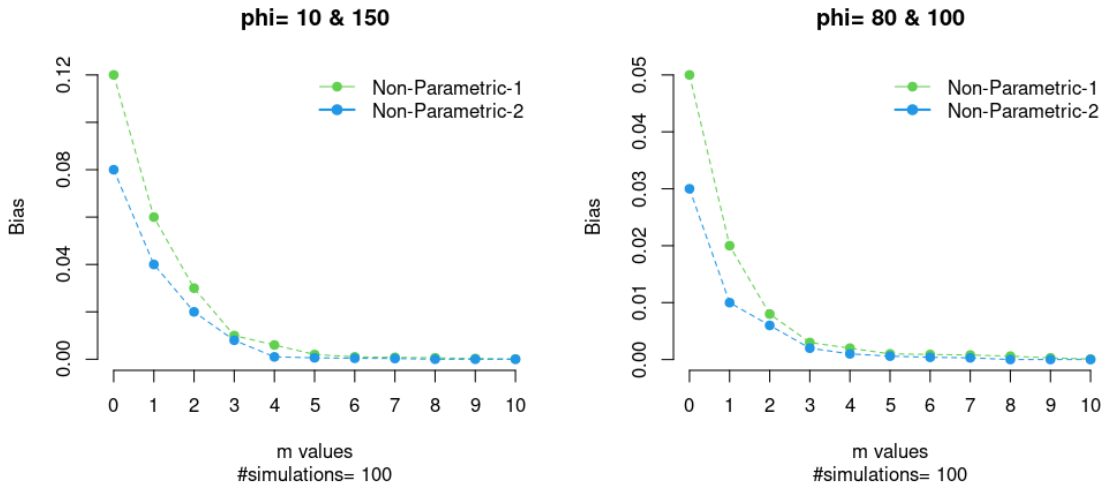

Figure 9. Bias between the estimated and the true covaraince functions vs different m values for both non-parametric approaches

It can be seen from the above graphs that the overall approximation between the estimated and true covariance functions from both approaches seems good for all m

values, with larger bias when m is small.

To investigate the performance of both approaches, we use the IPE value that was defined in Section 3.3.5. The smaller the IPE value is, the better performance an approach has. Further, we are interested in how both non-parametric approaches perform when the underlying process is axially symmetric and its special case, longitudinally reversible for all three mean structures. For each of the scenarios, we select several latitude points over the globe and compute their IPE values. The following plots illustrate how the IPE values change with latitudes from south to north of the globe for all three cases.

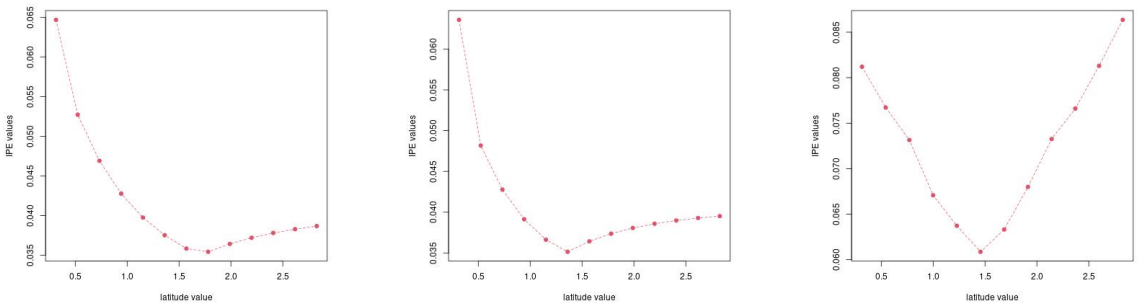

Figure 10. IPE values vs a range of latitudes for non-parametric approach 1 with longitudinally reversible processes under zero means (left plot), constant means (center plot), and different means (right plot).

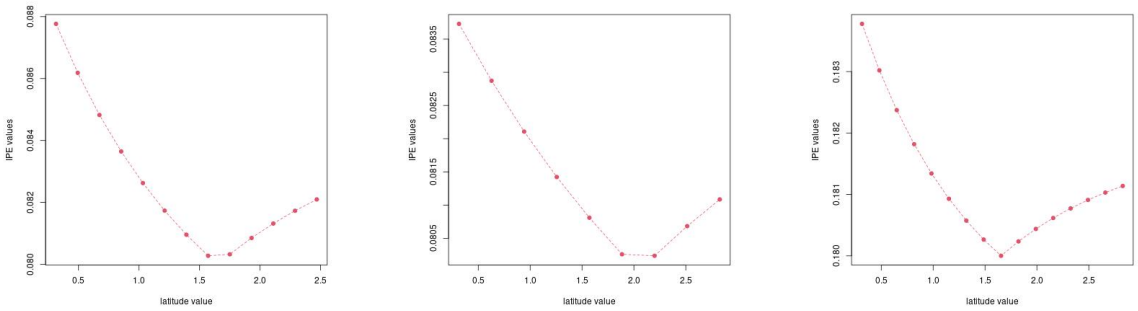

Figure 11. IPE values vs a range of latitudes for non-parametric approach 1 with axially symmetric processes under zero means (left plot), constant means (center plot), and different means (right plot).

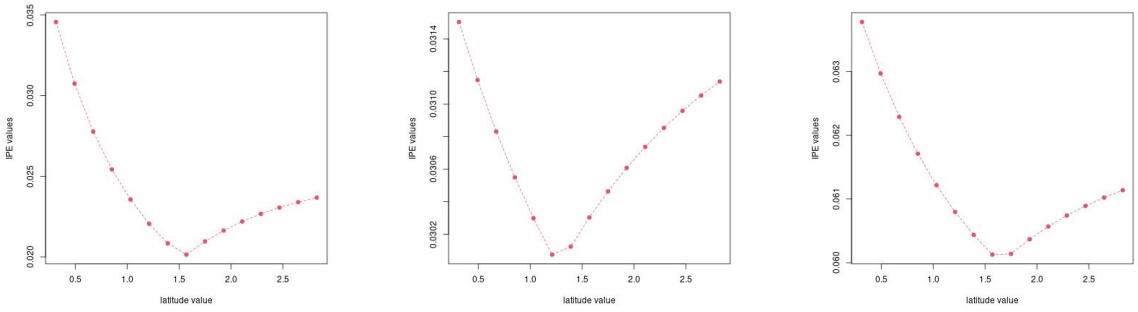

Figure 12. IPE values vs a range of latitudes for non-parametric approach 2 with longitudinally reversible processes under zero means (left plot), constant means (center plot), and different means (right plot).

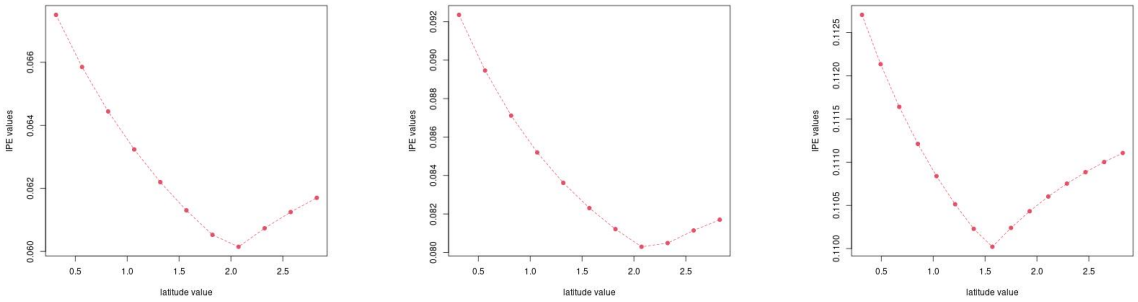

Figure 13. IPE values vs a range of latitudes for non-parametric approach 2 with axially symmetric processes under zero means (left plot), constant means (center plot), and different means (right plot).

Based on the above graphs, the smallest IPE values appear in the middle of latitudes, which indicates both approaches perform the best when the values are to be predicted are on the latitudes around the equator, compared with the prediction for latitudes near north and south poles. Overall, approach-2 has comparatively better performance than approach 1. To emphasize this more, we randomly selected one of the latitudes and performed kriging on that latitude. We computed the IPE values for both approaches over all scenarios. The following table displays the IPE values for both non-parametric approaches.

Table 7. IPE values for non-parametric approach 1 and 2 with all scenarios.

Process	Mean structure	Non-Para-1 (Mean(MSE))	Non-Para-2 (Mean(MSE))
Longitudinally reversible	Zero mean	1.143(0.324)	1.005(0.01)
	Constant mean	1.143(0.297)	1.005(0.01)
	Different mean	1.640(0.250)	1.006(0.01)
Axially symmetric	Zero mean	1.315(7.79)	1.305(1.215)
	Constant mean	1.274(5.99)	1.252(1.05)
	Different mean	1.305(1.215)	0.968(0.83)

According to the above table, due to smaller biases and MSE values, we can conclude that both approaches perform better for longitudinally reversible processes than axially symmetric processes under all three mean structures. This might be due to the fact that the covariance function for axially symmetric processes is complex, leading to the estimation with bigger errors. In addition, it seems that the approach 2 performs equally well compared to the approach 1.

Unlike parametric approaches, we noticed both approaches of non-parametric methods have smaller biases and MSEs for IPE values. Nevertheless, for axially symmetric processes, the non-parametric approach 1 results in slight differences compared to the non-parametric approach 2. We also investigate this further through box-plots for all 100 IPE values. Box plots are used to show overall patterns of response for a group.

They provide a useful way to visualise the range and other characteristics of responses for a large group.

The following two plots illustrate the distribution of IPE values under both axially symmetric processes on the non-parametric approaches. Based on the plots, we can conclude that very few values are far away from the central of the distribution. The medians (which are generally close to the average) are all at the same level in both approaches. In general, box plots are comparatively short, which indicates that overall IPE values have a high level of agreement from each other.

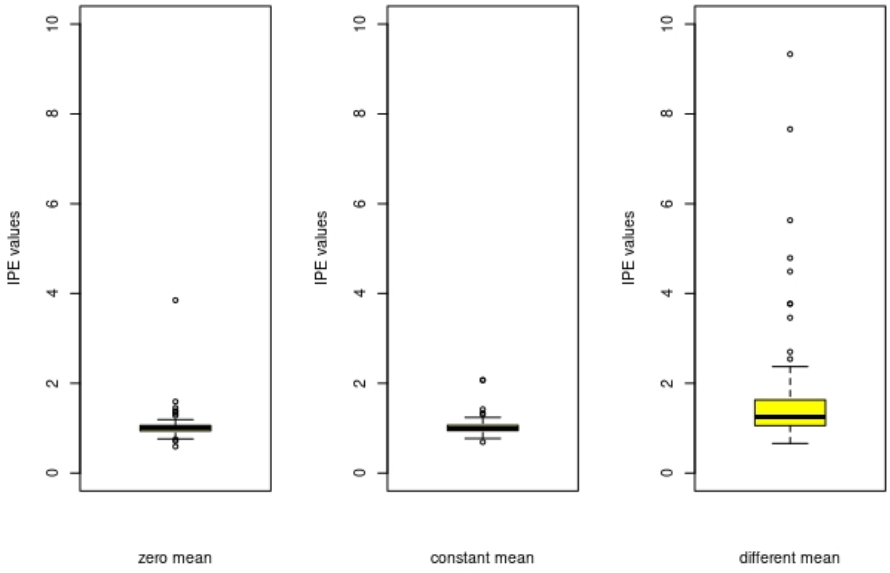

Figure 14. Box plots of IPE values for non-parametric 1 with axially symmetric processes under zero means (left plot), constant means (center plot), and different means (right plot).

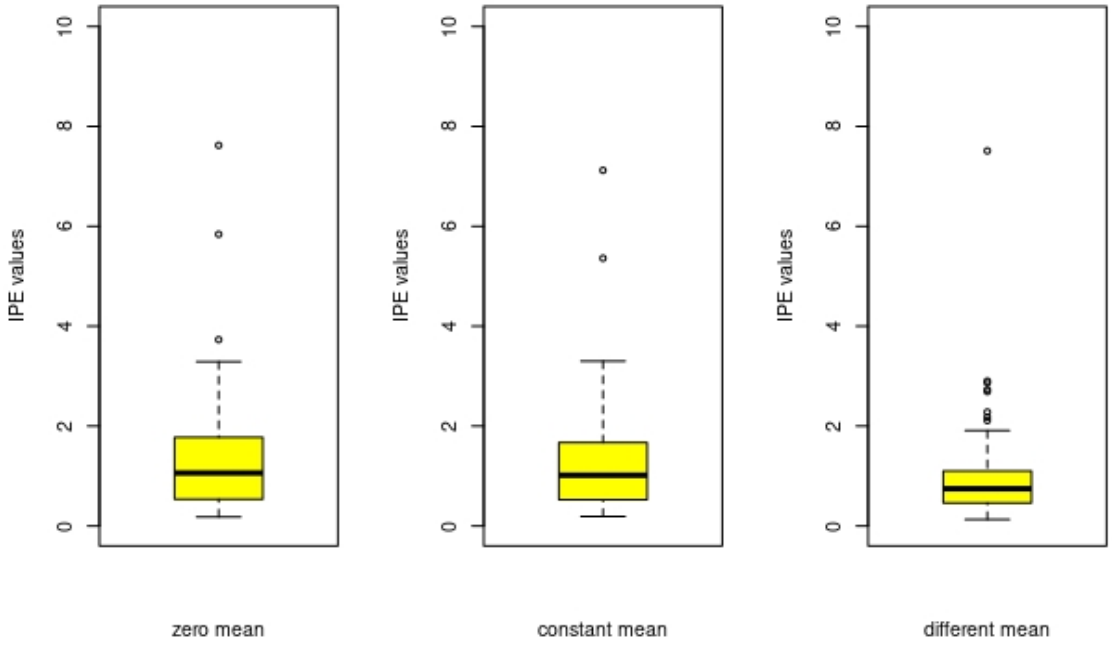

Figure 15. Boxplots of IPE values for non-parametric 2 with axially symmetric processes under zero means (left plot), constant means (center plot), and different means (right plot).

CHAPTER V

SUMMARY AND FUTURE WORK

.

5.1 Summary

In this dissertation, we considered spatial prediction when a random process is axially symmetric on the sphere. We first decompose an axially symmetric process as Fourier series on circles, where the Fourier random coefficients can be expressed as circularly-symmetric complex random processes. The estimation of covariance functions for complex random processes is then obtained through both parametric and non-parametric approaches, where the least squared error method and the Wavelet-Galerkin method are applied, respectively. Ordinary kriging is then conducted on possibly complex random fields and predicted data values are be computed through the inverse Discrete Fourier transformation. The performance of our proposed approaches were compared with simulation studies, under the assumption of axial symmetry and longitudinal reversibility with three mean structures: zero mean, constant mean, and different means on different latitudes. We conclude that all proposed kriging methods seem to perform well on all scenarios considered.

5.2 Future Research

There are a few areas that would extend this dissertation research. As we have noticed from Chapter IV, the LSE method seems to be a practical approach to estimate

the covariance function that is critical for kriging. Therefore, one would be very interested in investigating the asymptotic properties of the LSE estimator, including but not limited to the unbiasedness, consistency, and the asymptotic normality. Furthermore, we could also improve the LSE method with more parameters such as $C1\&p$. In addition, we noted that our simulation studies have been heavily depending on the covariance function that was proposed by [VWZ21]. There has been limited research on axially symmetric covariance models in literature, in particular, the construction of axially symmetric covariance models on S^2 that could be practically useful in analyzing massive global data sets as well as the implementation of our proposed kriging methods. Moreover, with the complexity and dimensionality of global data, it is very important that practically useful parametric models with interpretable parameters are available for geography and environmental scientists.

Finally, we note that the unbiased estimator of $R(\phi_P, \phi_Q, \Delta\lambda)$ is critical for our non-parametric approaches, where we further assume that we have *i.i.d.* copies of gridded data structures. Such an assumption might not be feasible in practice, and hence it needs to be generalized. For example, we may consider that we have a spatio-temporal sequence of gridded data structures, and we could replace the *i.i.d.* assumption. For example, an AR(1) model, that is, the sequence of gridded data structures might be dependent, but following an auto-regression model with an order of 1.

REFERENCES

[Ada17] Adaramola, B.O., *Circulant matrices on global data analysis*. Master's thesis, Department of Mathematics and Statistics, University of North Carolina at Greensboro, 2017.

[Arachchige21] Arachchige R.R.T.A.L., *The Wavelet-Galerkin method on global random process*. Ph.D Dissertation, UNC Greensboro, NC, 2021.

[BISHOP20] Bishop M.P., B W. Young, Da Huo, Zhaohui Chi., *Spatial Analysis and Modeling in Geomorphology*. 2020.

[Brock91] Peter J. Brockwell, Richard A. Davis., *Time Series: Theory and Methods,* 1991.

[BUS20] Bussberg, N,. *Environmental applications of temporal, spatial, and spatial-temporal statistics*. PhD Dissertation, Indiana University, Bloomington, IN, 2020.

[CHR88] Christopher B., David W.Fox., *Schur complements and the Weinstein-Aronszajn theory for modified matrix eigenvalue problems*. Linear Algebra and its Applications, Vol-108, 1988.

[Chu07] Chua, David K, Eric C, Mark, *Network Kriging.* Selected Areas in Communications, IEEE Journal 24, 2263-2272, 2007.

[Cre88] Cressie, N.A.C., *Spatial prediction and ordinary kriging.* Math Geol 20, 405–421, 1988.

[Cre93] Cressie, N.A.C., *Statistics for spatial data*. Wiley-Interscience, 1993.

[Deu98] Deutsch CV, Journel AG, *GSLIB: geostatistical software library and user's guide.* Applied geostatistics series, 2nd edn. Oxford University Press, New York 1998.

[Ger20] Gerassis, Saki B., Carlos A, Ribeiro M., Maria A, A., Taboada, J., *Mapping occupational health risk factors in the primary sector—A novel supervised machine learning and Area-to-Point Poisson kriging approach.* Spatial Statistics, 42, 100-434, 2020.

[Good08] Goodchild, Michael F., *Encyclopedia of GIS*. Springer US,Boston, MA, 200-203, 2008.

[Grib20] Gribov, Alexander,Krivoruchko, Konstantin, *Empirical Bayesian kriging implementation and usage.* Science of The Total Environment, 2020.

[Hes20] Hesamian G., Akbari M.G., *A kriging method for fuzzy spatial data.* International Journal of Systems Science, 1945-1958, 2020.

[HZR12] Huang, C., Zhang, H., Robeson, S. M., *A simplified representation of the covariance structure of axially symmetric processes on the sphere*. Statistics and Probability Letters 82(7), 1346-1351, 2012.

[HZRS19] Huang, C., Zhang, H., Robeson, S. M., Shields, J., *Intrinsic Random Functions on the Sphere*. Statistics and Probability Letters, 146, 7-14, 2019.

[Jae2014] Jaehong J, Mikyoung J., *A class of Matérn-like covariance functions for smooth processes on a sphere.* Spatial Statistics, 11 ,1-18,2015.

[Jon63] Jones, R.H., *Stochastic processes on a sphere*. The Annals of Mathematical Statistics, Vol. 2, 213-218, 1963.

[JS08] Jun, M., Steal M.L., *Nonstionary covariance models for global data*. The Annals of Applied Statistics, Vol. 2, No. 4, 1271–1289, 2008.

[Krig] Daniel G. Krige, *A statistical approach to some basic mine valuation problems on the Witwatersrand.* Journal of The South African Institute of Mining and Metallurgy, 201-203, 1951.

[Li20] Li, Yingchang Li, Mingyang Liu, Zhenzhen Li, Yingchang., *Combining Kriging Interpolation to Improve the Accuracy of Forest Above ground Biomass Estimation Using Remote Sensing Data.* IEEE Access, 2020.

[Math63] Mathéron, G., *Principles of geostatistics.* Economic Geology 58, 1246–1266, 1963.

[Math71] Mathéron, G, *La théorie des fonctions aléatoires intrinsèques généralisées.* Note géostatistique N° 117, Technical report N-252, Centre de Géostatistique, Fontainebleau, 1971.

[Ord83] Ord, J. K., *Kriging entry in encyclopedia of statistical sciences.* Vol. 4, pp. 411–413, New York: Wiley Publication, 1983.

[PAC18] Porcu, E. , Algeria, A., Crippa, P., *Axially symmetric models fo global data: A journey between geostatistics and stochastic generators*. Enviromentrices, 30 (1), 2018.

[PHQ02] Phoon, K. K., Huang, S. T., Quek, S. T., *Implementation of Karhunen-Loeve expansion for simulation using a Wavelet-Galerkin scheme*. Probabilistic Engineering Mechanics 17 293-303, 2002.

[RW05] Rasmussen, C.E., Williams, C.K.I., *Gaussian Processes for machine learning*. The MIT Press, Cambridge, Massachusetts, 2005.

[Sand15] Sandra De Iaco, Donato Posa., *Wind velocity prediction through complex kriging: formalism and computational aspects.* Environmental Ecology Statistics, 23, 115-139, 2016.

[Sato19] Sato, Koya Inage, Kei Fujii, Takeo., *On the Performance of Neural Network Residual Kriging in Radio Environment Mapping.* IEEE Access, 7, 94557-94568, 2019.

[Seti20] Setiyoko, Andie, Basaruddin, Arymurthy T, Aniati., *Minimax Approach for Semivariogram Fitting in Ordinary Kriging.* IEEE Access, 2020.

[Shao20] Shao, Yanchuan Ma, Zongwei Wang, Jianghao Bi, Jun., *Estimating daily ground-level PM2.5 in China with random-forest-based spatiotemporal kriging.* Science of The Total Environment, 2020.

[Ste99] Stein, M.L., *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.

[Ste07] Stein, M.L., *Spatial variation of total column ozone on a global scale*. Annals of Applied Statistics, 191-210, 2007.

[Tuo20] Tuo R, Wang W., *Kriging Prediction with Isotropic Matern Correlations: Robustness and Experimental Designs.* J. Mach. Learn. Res., 2020.

[Van16] Vanlengenberg, C.D., *Data generation and estimation for axially symmetric processes on the sphere*. Ph.D Dissertation. UNC Greensboro, NC, 2016.

[VWZ21] Vanlengenberg, C.D., Wang, W., Zhang, H., *Data generation for axially symmetric processes on the sphere*. Communication in Statistics-Simulation and Computation, 50, 1750 - 1769, 2019.

[UNWIN09] Unwin D.J., *International Encyclopedia of Human Geography,* 2009.

[Wack03] Wackernagel, H., *Multivariate Geostatistics: An Introduction with Applications.* Springer-Verlag, Berlin, 2003.

[Wad06] Wade, Sommer T., *An illustrated dictionary of geographic information systems*. ESRI Press, Redlands, California, USA, 2006.

[Woj18] Wojciechm M., *Kriging method optimization for the process of DTM creation based on huge data sets obtained from MBESs.* Geosciences, 8(12), 2018.

[Yan13] Yang, L., *Non-parametric and semi-parametric estimation of spatial covariance function*. Ph.D Dissertation. Iowa State University, 2013.

[Zhu16] Yang Li, Zhengyuan Zhu., *Modeling non-stationary covariance function with convolution on sphere.* Computational Statistics & Data Analysis, 104, 233-246, 2016.

[Zhu18] A-Xing Zhu, Guonian Lu, Jing Liu, Cheng-Zhi Qin, Chenghu Zhou., *Spatial prediction based on Third Law of Geography*. 225-240,Annals of GIS,2018.