Data Generation and Estimation for Axially Symmetric Processes on the Sphere

Chris Vanlangenberg[1]
Advisor: Dr. Haimeng Zhang[2]

University of North Carolina at Greensboro

September 25, 2016

[1]cdvanlan@uncg.edu
[2]h_zhang5@uncg.edu

Abstract

Review and update later

Global-scale processes and phenomena are of utmost importance in the geophysical sciences. Data from global networks and satellite sensors have been used to monitor a wide array of processes and variables, such as temperature, precipitation, etc. In this dissertation, we are planning to achieve explicitly the following objectives,

1. Develop both non-parametric and parametric approaches to model global data dependency.

2. Generate global data based on given covariance structure.

3. Develop kriging methods for global prediction.

4. Explore one or more of the popularly discussed global data sets in literature such as MSU (Microwave Sounding Units) data, the tropospheric temperature data from National Oceanic and Atmospheric and TOMS (Total Ozone Mapping Spectrometer) data, total column ozone from the Laboratory for Atmospheres at NASA's Goddard Space Flight Center Administration satellite-based Microwave Sounding Unit.

Global scale data have been widely studied in literature. A common assumption on describing global dependency is the second order stationarity. However, with the scale of the Earth, this assumption is in fact unrealistic. In recent years, researchers have focused on studying the so-called axially symmetric processes on the sphere, whose spatial dependency often exhibit homogeneity on each latitude, but not across the latitudes due to the geophysical nature of the Earth. In this research, we have obtained some results on the method of non-parametric estimation procedure, in particular, the method of moments, in the estimation of spatial dependency. Our initial result shows that the spatial dependency of axially symmetric processes exhibits both anti-symmetric and symmetric characteristics across latitudes. We will also discuss detailed methods on generating global data and finally we will outline our methodologies on kriging techniques to make global prediction.

Chapter 1

Introduction

In this chapter we will give a brief introduction to some of the basic concepts in spatial statistics, which are necessary to follow for the rest of this dissertation. More specifically, we will have discuss about stationarity and intrinsic stationarity, covariance and variogram functions and their properties, mean square continuity and differentiability, spectral representations and spectral densities, complex random processes and Gaussian random vectors, as well as some basic properties related to circulant and circulant block matrices. Finally we will give an outline of this dissertation at the last section.

1.1 Spatial random field

A random process is a collection of random variables $\{Z(s) : s \in D\}$, defined in a common probability space, that take values on a specific domain D. Generally, D may take a variety of forms as given below.

- $s \in D = N$: $Z(s)$ is a random sequence which is used in time series.
- $s \in D = R^1$: $Z(s)$ is a random process which is also referred as a stochastic process.
- $s \in D = R^d$: $Z(s)$ is a random filed or a spatial process if $d > 1$
- $s \in D = S^2$: $Z(s)$ is a random process on the sphere.
- $s \in D = R^d \times R$: $Z(s)$ is a spatio-temporal process which involves location and time.

A real-valued spatial process in d-dimensional Euclidean space R^d or a spatial random field can be denoted as $\{Z(x) : x \in D \subset \mathbb{R}^d\}$ where x is the location, varying over a fixed domain D. The distribution of the random process $Z(s)$ is characterized by its finite-dimensional distribution function, that is, the distribution function of the random vector $Z(\underset{\sim}{X}) = (Z(x_1), \ldots, Z(x_n))$ given by

$$F\{Z(h_1), \ldots, Z(h_n)\} = P\{Z(x_1) \leq h_1, \ldots, Z(x_n) \leq h_n\}, \tag{1.1.1}$$

for any n and any sequence of locations $(x_1, x_2, \ldots, x_n)$.

1.1.1 Stationarity and Isotropy

A spatial random field $Z(x)$ is said to be strictly stationary, if for all finite $n,\ x_1, \ldots, x_n \in \mathbb{R}^d$, $h_1, \ldots, h_n \in \mathbb{R}$ and $x \in \mathbb{R}^d$, $Z(x)$ is invariant under translation, that is,

$$P\{Z(x_1 + x) \leq h_1, \ldots, Z(x_n + x) \leq h_n\} = P\{Z(x_1) \leq h_1, \ldots, Z(x_n) \leq h_n\} \tag{1.1.2}$$

Strict stationarity is normally too strong a condition as it involves the distribution of the random field. Another commonly used but weaker assumption is the weak stationarity. More specifically, a random process $Z(x)$ is weakly stationary if

$$\begin{aligned} E(Z(x)) &= \mu \\ E(Z^2(x)) &< \infty \\ C(h) &= Cov(Z(x), Z(x+h)) \end{aligned} \tag{1.1.3}$$

In other words a random process $Z(x)$ is weakly stationary (or simply stationarity throughout the rest of this dissertation) if it has constant mean and finite second moment and its (auto-)covariance function $C(h)$ solely depends on the spatial distance of two locations. Further, a strictly stationary random field with finite second moment is weakly stationary, but weak stationarity does not imply strict stationarity unless $Z(x)$ is a Gaussian random field, under which both stationarity are equivalent, as the finite-dimensional distribution of a Gaussian random field is multivariate normal and is uniquely determined by the first and second moments.

The auto-covariance function $C(h)$ of a stationary process $Z(x)$ on $\mathbb{R}^d$ has the following properties.

(i) $C(0) \geq 0$;

(ii) $C(h) = C(-h)$;

(iii) $|C(h)| \leq C(0)$;

(iv) If $C_1(h), C_2(h), \ldots, C_n(h)$ are valid covariance functions then each of the following functions $C(h)$ is also a valid covariance function.

   (a) $C(h) = a_1C_1(h) + a_2C_2(h)$, $\forall a_1, a_2 \geq 0$;

   (b) $C(h) = C_1(h)C_2(h)$;

   (c) $\lim_{n\to\infty} C_n(h) = C(h),\ \forall h \in \mathbb{R}^d$.

A function $C(\cdot)$ on $\mathbb{R}^d$ is non-negative definite if and only if

$$\sum_{i,j=1}^{N} a_i a_j C(x_i - x_j) \geq 0, \tag{1.1.4}$$

for any integer N, any constants $a_1, a_2, \ldots, a_N$, and any locations $x_1, x_2, \ldots, x_N \in \mathbb{R}^d$. Positive definiteness is a necessary and sufficient condition such that a function is a valid covariance function.

A weakly stationary process with a covariance function $C(||h||)$ which is free from direction is called isotropy. The random field, $Z(x)$, on $\mathbb{R}^d$ is strictly isotropy if the joint distributions are invariant under all rigid motions. *i.e.,* for any orthogonal $d \times d$ matrix H and any $x \in \mathbb{R}^d$

$$P\{Z(Hx_1 + x) \leq h_1, \ldots, Z(Hx_n + x) \leq h_n\} = P\{Z(x_1) \leq h_1, \ldots, Z(x_n) \leq h_n\} \quad (1.1.5)$$

Isotropy assumes that it is not required to distinguish one direction from another for the random field $Z(x)$.

Variogram function is an alternative to the covariance function proposed by Matheron (1973). It is defined as the variance of the difference between random fields at two locations, that is

$$2\gamma(h) = Var(Z(s+h) - Z(x)). \quad (1.1.6)$$

Here $\gamma(h)$ is called the semivariogram. If the variogram function solely depends on the distance of the two locations, then the process with finite constant mean is said to be intrinsically stationary. If $Z(x)$ is further assumed to be stationary with covariance function $C(h)$, then $\gamma(h) = C(0) - C(h)$. Intrinsic stationarity is defined in terms of variogram and it is more general than (weak) stationarity that is defined in terms of covariance. Clearly, when $C(h)$ is known, we can obtain $\gamma(h)$ but the reverse is not true. For example we consider a linear semivariogram function given below,

$$\gamma(h) = \begin{cases} a^2 + \sigma^2 h & h > 0 \\ 0 & otherwise \end{cases}$$

when $\lim_{h \to \infty} \gamma(h) \to \infty$ thus the process with the above semivariogram is not weakly stationary and $C(h)$ does not exist.

Parallel to the positive definiteness about the covariance function, the variogram is conditionally negative definite, that is,

$$\sum_{i,j=1}^{N} a_i a_j 2\gamma(x_i - x_j) \leq 0, \quad (1.1.7)$$

for any integer N, any constants $a_1, a_2, \ldots, a_N$ with $\sum_{i=1}^{N} a_i = 0$, and any locations $x_1, x_2, \ldots, x_N \in \mathbb{R}^d$.

1.1.2 Mean square continuity & differentiability

There is no simple relationship between $C(h)$ and the smoothness of $Z(x)$. For a sequence of random variables $X_1, X_2, \ldots$ and a random variable X defined on a common probability space. Define $X_n \xrightarrow{L^2} X$ if $E(X^2) < \infty$ and $E(X_n - X)^2 \to 0$ as $n \to \infty$. We then say $\{X_n\}$ converges in L^2 if there exists such a X.

Suppose $Z(x)$ is a random field on $\mathbb{R}^d$, then $Z(x)$ is mean square continuous at x if

$$\lim_{h \to 0} E(Z(x+h) - Z(x))^2 = 0.$$

If $Z(x)$ is stationary and $C(\cdot)$ is the covariance function then $E(Z(x+h)-Z(x))^2 = 2(C(0) - C(h))$. Therefore $Z(x)$ is mean square continuous if and only if $C(\cdot)$ is continuous at the origin.

1.1.3 Spectral representation of a random field

Suppose $\omega_1, \ldots, \omega_n \in \mathbb{R}^d$ and let $Z_1, \ldots, Z_n$ be mean zero complex random variables with $E(Z_i \bar{Z}_j) = 0, i \neq j$ *and* $E|Z_i|^2 = f_i$. Then the random sum

$$Z(x) = \sum_{k=1}^{n} Z_k e^{i\omega_k^T x}. \tag{1.1.8}$$

is a weakly stationary complex random field in $\mathbb{R}^d$ with covariance function $C(x) = \sum_{k=1}^{n} f_k e^{i\omega_k^T x}$

Further, if we think about the integral as a limit in L^2 of the above random sum, then the covariance function can be represented as,

$$C(x) = \int_{\mathbb{R}^d} e^{i\omega^T x} F(d\omega) \tag{1.1.9}$$

where F is the so-called spectral distribution. There is a more general result from Bochner.

Theorem 1.1.1 (Bochner's Theorem)

A complex valued covariance function $C(\cdot)$ on $\mathbb{R}$ for a weakly stationary mean square continuous complex-valued random field on $\mathbb{R}^d$ if and only if it can be represented as above, where F is a positive measure.

If F has a density (spectral density, denoted by f) with respect to Lebesgue measure, (*i.e.* if such f exists) we can use the inversion formula to obtain f

$$f(\omega) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\omega^T x} C(x) dx \tag{1.1.10}$$

1.1.4 Spectral densities

Here we provide some examples of isotropic covariance functions and their corresponding spectral densities.

(i) Rational Functions that are even, non-negative and integrable the corresponding covariance functions can be expressed in terms of elementary functions. For example if $f(\omega) = \phi(\alpha^2 + \omega^2)^{-1}$, then $C(h) = \pi\phi\alpha^{-1}e^{-\alpha|h|}$ (obtained by contour integration).

(ii) Gaussian are the most commonly used covariance function for a smooth process on $\mathbb{R}$ where the covariance function is given by $C(h) = ce^{-\alpha h^2}$ and the corresponding spectral density is $f(\omega) = \frac{1}{2\sqrt{\pi\alpha}}ce^{\frac{-\omega^2}{4\alpha}}$.

(iii) *Matérn* class has more practical use and more frequently used in spatial statistics. The spectral density of the form $f(\omega) = \frac{1}{\phi(\alpha^2+\omega^2)^{\nu+1/2}}$ where $\phi, \nu, \alpha > 0$ and the corresponding covariance function given by,

$$C(h) = \frac{\pi^{1/2}\phi}{2^{\nu-1}\Gamma(\nu + 1/2)\alpha^{2\nu}}(\alpha|h|)^{\nu}Y_{\nu}(\alpha|t|) \tag{1.1.11}$$

where Y_ν is the modified Bessel function, the larger the ν smoother the Y. Further, Y will be m times square differentiable iff $\nu > m$. When ν is in the form of $m + 1/2$ with m a non negative integer. The spectral density is rational and the covariance function is in the form of $e^{-\alpha|h|}\cdot$ polynomial($|h|$) for example, when $\nu = \frac{1}{2}$ $C(h)$ corresponds to exponential model and $\nu = \frac{3}{2}$ is transformation of exponential family of order 2.

$$\begin{aligned} \nu = 1/2 \quad &: \quad C(h) = \pi\phi\alpha^{-1}e^{-\alpha|h|} \\ \nu = 3/2 \quad &: \quad C(h) = \frac{1}{2}\pi\phi\alpha^{-3}e^{-\alpha|h|}(1 + \alpha|h|) \end{aligned}$$

1.2 Circularly-symmetric Gaussian random vectors

Let $\underset{\sim}{Z} = (Z_1, Z_2, \ldots, Z_n)^T$, where $Z_j = (Z_j^{Re}, Z_j^{Im})^T$ and $j = 1, 2, \ldots, n$ be a zero mean $2n$ complex random vector of dimension $2n$. Then its covariance matrix K_Z and the pseudo-covariance matrix M_Z are defined as follow.

$$K_{\underset{\sim}{Z}} = E[\underset{\sim}{Z}\underset{\sim}{Z}^*] \tag{1.2.1}$$

$$M_{\underset{\sim}{Z}} = E[\underset{\sim}{Z}\underset{\sim}{Z}^T] \tag{1.2.2}$$

where $\underset{\sim}{Z}^*$ is the conjugate transpose of $\underset{\sim}{Z}$.

Generally, to characterize the relationship of a complex random vector, one needs both covariance and pseudo-covariance matrices. First note that a complex random variable

$Z = Z^{Re} + iZ^{Im}$ is (complex) Gaussian, if Z^{Re}, Z^{Im} both are real and they are jointly Gaussian. Now we consider a vector $\underset{\sim}{Z} = (Z_1, Z_2)^T$ where $Z_1 = Z_1^{Re} + iZ_1^{Im}$ and $Z_2 = Z_1^*$ $(Z_2^{Re} = Z_1^{Re}, Z_2^{Im} = -Z_1^{Im})$. The four real and imaginary parts of $\underset{\sim}{Z}$ are jointly Gaussian (each follows $N(0, 1/2)$) (so $\underset{\sim}{Z}$ is complex Gaussian).

The covariance and pseudo-covariance matrices defined by 1.2.1 and 1.2.2, respectively, are given by

$$M_Z = E\begin{bmatrix} Z_1^2 & Z_1Z_1^* \\ Z_1Z_1^* & Z_1^2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and

$$K_Z = E\begin{bmatrix} Z_1Z_1^* & Z_1^2 \\ Z_1^2 & Z_1Z_1^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

It is easy to note that $E[Z_1^2] = E[Z_1^{Re}Z_1^{Re} - Z_1^{Im}Z_1^{Im}] = 1/2 - 1/2 = 0$. If both Z_1 and Z_2 are real, then covariance and pseudo-covariance matrices are the same, *i.e.*, $M_Z \equiv K_Z$

The covariance matrix of real $2n$ random vector $\underset{\sim}{Z} = (\underset{\sim}{Z}^{Re}, \underset{\sim}{Z}^{Im})^T$, where $\underset{\sim}{Z}^{Re} = (Z_1^{Re}, Z_1^{Re}, \ldots, Z_1^{Re})$ and $\underset{\sim}{Z}^{Im} = (Z_1^{Im}, Z_2^{Im}, \ldots, Z_n^{Im})$ can be determined by both $K_{\underset{\sim}{Z}}$ and $M_{\underset{\sim}{Z}}$ given as follow.

$$\begin{aligned} E[\underset{\sim}{Z}^{Re}\underset{\sim}{Z}^{Re}] &= \frac{1}{2}Re(K_{\underset{\sim}{Z}} + M_{\underset{\sim}{Z}}), \\ E[\underset{\sim}{Z}^{Im}\underset{\sim}{Z}^{Im}] &= \frac{1}{2}Re(K_{\underset{\sim}{Z}} - M_{\underset{\sim}{Z}}), \\ E[\underset{\sim}{Z}^{Re}\underset{\sim}{Z}^{Im}] &= \frac{1}{2}Im(-K_{\underset{\sim}{Z}} + M_{\underset{\sim}{Z}}), \\ E[\underset{\sim}{Z}^{Im}\underset{\sim}{Z}^{Re}] &= \frac{1}{2}Im(K_{\underset{\sim}{Z}} + M_{\underset{\sim}{Z}}) \end{aligned} \tag{1.2.3}$$

We can get the covariance of $\underset{\sim}{Z} = (\underset{\sim}{Z}^{Re}, \underset{\sim}{Z}^{Im})^T$ as follows,

$$\begin{aligned} Cov(\underset{\sim}{Z}) &= E(\underset{\sim}{Z}\underset{\sim}{Z}^T) \\ &= \begin{pmatrix} E[\underset{\sim}{Z}^{Re}\underset{\sim}{Z}^{Re}] & E[\underset{\sim}{Z}^{Re}\underset{\sim}{Z}^{Im}] \\ E[\underset{\sim}{Z}^{Im}\underset{\sim}{Z}^{Re}] & E[\underset{\sim}{Z}^{Im}\underset{\sim}{Z}^{Im}] \end{pmatrix} \end{aligned}$$

Now we introduce circularly-symmetric random variables and vectors. A complex random variable Z is circularly-symmetric if both Z and $e^{i\phi}Z$ have the same probability distribution for all real ϕ. Since $E[e^{i\phi}Z] = e^{i\phi}E[Z]$, any circularly-symmetric complex random variable must have $E[Z] = 0$, in other words its mean must be zero.

For a circularly-symmetric complex random vector, we have the following theorem [Gallager, 2008].

Theorem 1.2.1 (Gallager, 2008)
Let $\underset{\sim}{Z}$ be a zero mean Gaussian random vector then $M_{\underset{\sim}{Z}} = 0$ if and only if $\underset{\sim}{Z}$ is circularly-symmetric.

In spatial statistics, sometimes it is more convenient to use complex valued random functions, rather than real valued random functions. We say, $Z(x) = U(x) + iV(x)$ is a complex random field if $U(x), V(x)$ are real random fields. If $U(x), V(x)$ are stationary so does $Z(x)$. The covariance function can be defined as,

$$C(h) = cov(Z(x+h), \overline{Z(x)}), \quad C(-h) = \overline{C(h)}.$$

For any complex constants $c_1, \ldots, c_n$, and any locations $x_1, x_2, \ldots, x_n$,

$$\sum_{i,j=1}^{n} c_i \bar{c}_j C(x_i - x_j) \geq 0 \tag{1.2.4}$$

1.3 Circulant matrix

A square matrix $A_{n\times n}$ is a circulant matrix if the elements of each row (except first row) has the previous row shifted by one place to the right.

$$A = circ[a_o, a_1, \cdots, a_{n-1}] \;=\; \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \cdots & a_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix}. \tag{1.3.1}$$

The eigenvalues of A are given by

$$\begin{aligned} \lambda_l &= \sum_{k=0}^{n-1} a_k e^{i2lk\pi/n} \\ &= \sum_{k=0}^{n-1} a_k \rho_l^k, \qquad l = 0, 1, 2, \cdots, n-1, \end{aligned}$$

where $\rho_l = e^{i2\pi l/n}$ represents the lth root of 1, and the corresponding (unitary) eigenvector is given by

$$\psi_l = \frac{1}{\sqrt{n}}(1, \rho_l, \rho_l^2, \cdots, \rho_l^{n-1})^T.$$

If matrix A is real symmetric, that is, $a_i = a_{n-i}$, then its eigenvalues are real. More specifically, for even $n = 2N$ the eigenvalues $\lambda_j = \lambda_{n-j}$ or there are either two eigenvalues or none with odd multiplicity, for odd $n = 2N-1$ the eigenvalue λ_0 equal to any λ_j for $1 \leq j \leq N-1$ or λ_0 occurs with odd multiplicity. A square matrix B is Hermitian, if and only if $B^* = B$ where B^* is the complex conjugate. If B is real then $B^* = B^T$. According to Tee (2005) Hermitian matrices has a full set of orthogonal eigenvectors with corresponding real eigenvalues.

1.3.1 Block circulant matrices

The idea of a block circulant matrix was first proposed by Muir (1920). A matrix $B_{np\times np}$ is a block-circulant matrix if it has the following form,

$$B = bcirc[A_o, A_1, \cdots, A_{n-1}] \;=\; \begin{bmatrix} A_0 & A_1 & A_2 & \cdots & A_{n-1} \\ A_{n-1} & A_0 & A_1 & \cdots & A_{n-2} \\ A_{n-2} & A_{n-1} & A_0 & \cdots & A_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_1 & A_2 & A_3 & \cdots & A_0 \end{bmatrix}. \tag{1.3.2}$$

where A_j are $(p \times p)$ sub-matrices of complex or real valued elements. De Mazancourt and Gerlic (1983) proposed some methodologies to find the inverse of B. Let M be a block-permutation matrix

$$M = \begin{bmatrix} 0 & I_p & 0 & \cdots & 0 \\ 0 & 0 & I_p & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_p \\ I_p & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

where I_p is $p \times p$ identity matrix and B can be defined as follows,

$$B = \sum_{k=0}^{n-1} A_k M^k.$$

Define M^0 as $(np \times np)$ identity matrix and the eigenvalues of M given by ρ_l, the eigen-matrix of M can be given by $Q_{np\times np} = \{\underset{\sim}{\psi_0}, \underset{\sim}{\psi_1}, \ldots, \underset{\sim}{\psi_{n-1}}\}$. From Trapp (1973) it can be shown that $Q^{-1} = Q^*/n$ where Q^* is the conjugate transpose of Q now we can write,

$$M = QDQ^{-1} = \frac{QDQ^*}{n}$$

where D is a diagonal matrix and the diagonal elements $D_i \quad i = 0, 1, n-1$ are the discrete Fourier transform of the blocks A_j,

$$D_i = \sum_{k=0}^{n-1} A_k e^{i2lk\pi/n}$$

That is the inverse of matrix B takes the following form,

$$B^{-1} = Q \cdot \begin{pmatrix} D_0^{-1} & 0 & \cdots & 0 \\ 0 & D_1^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_n{-1} \end{pmatrix} \cdot Q^{-1}.$$

The eigenmatrix Q is solely depending on the dimension of B and the eigenvalues of B (ρ_l's) or in other words B is not depending on the blocks (A_j's), *i.e.* for any block diagonal matrix $D_{np\times np}$, QDQ^{-1} is a block circulant matrix and immediately follows that the inverse of the matrix B is also a block circulant matrix.

When $A_{j_{1\times 1}}$, $B = A$, $D_i^{-1} = \lambda^{-1}$, and the eigenmatrix has a dimension of $n \times n$ then

$$A^{-1} = Q\Lambda^{-1}Q^T \quad \text{where } \Lambda = \{\lambda_0, \ldots, \lambda_{n-1}\}$$

When A is real symmetric Q is real also symmetric and $Q^{-1} = Q^T$.

*We need to add in the special case when A_j's are symmetric (*by Tee, add citation*)*

1.4 The outline of this dissertation

In Chapter 2, we explore some of the properties of commonly used covariance and variogram estimators on the circle based on method of moments. Contrasting to the results given in time series and the Euclidean space, the MOM covariance estimator is biased and possibly nonidentificable due to the unestimable bias. On the other hand, the MOM variogram estimator is unbiased, but it can be shown to be inconsistent. Chapter 3 first introduces the random process on the sphere. We then discuss the homogeneous process and the spectral representation for its covariance function on the sphere. Our main focus on this chapter is the axially symmetric process and its covariance function representation, through the discrete fourier transform. The parametric models for characterizing such processes are also discussed. In particular, we extend the models given in Huang et al. (2012) and provide some graphical properties of those models. These extended models will be fully implemented in Chapter 4 for axially symmetric data generation. In Chapter 4, we explore the result given in Huang et al. (2012) to implement an algorithm for axially symmetric data generation. In particular, we observe that *... (more to come)*

In general for covariance function defined on a sphere (Stein (2007)) requires triple summation and required to estimate $O(n^3)$ parameters. In contrast, the covariance function defined by Huang et al. (2012) requires to estimate $O(n^2)$ parameters which is a huge reduction of computational compelextity and we will continue to use this covariance models in our approach on global data generation which is discussed in chapter 4.

. Finally, Chapter 5 gives a summary of this research and provides some further research directions.

Chapter 2

Literature Review

2.1 Spatial Data

What does it mean by spatial data? In general, spatial data or in other words geospatial data is information about a physical object or a measurement that can be represented by numerical values in a geographic coordinate system. As Cressie (1993) pointed out spatial data appeared to be in the form of maps in 1686 and spatial modeling paper in 1907. There are many questions that geoscientists and engineers are interested about spatial data. Many questions naturally arise such as how to model a spatial process and then use the model to make predictions about unobserved locations. There are many challenges when modeling spatial data; every point (location observed) is a random variable and only one observation/measurement is available. However, the number of unknowns to estimate are quite large compared to the available data, which is definitely a high-dimensional problem. As an example, if data were observed at 10 locations, one is estimating the variance-covariance matrix to characterize the spatial dependency for future predictions. Then there will have 55 unknown entities in the variance-covariance matrix to be estimated.

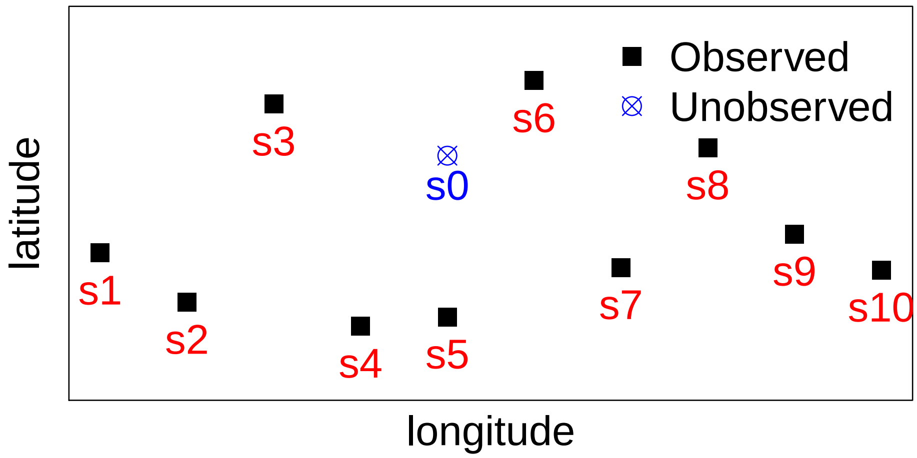

Figure 2.1: Some arbitary saptial data at 10 random locations

2.1.1 MSU data

Since 1978 Microwave Sounding Units (MSU) measure radiation emitted by the earth's atmosphere from NOAA polar orbiting satellites. The different channels of the MSU measure different frequencies of radiation proportional to the temperature of broad vertical layers of the atmosphere. Tropospheric and lower stratospheric temperature data are collected by NOAA's TIROS-N polar-orbiting satellites and adjusted for time-dependent biases by the Global Hydrology and Climate Center at the University of Alabama in Huntsville (UAH)[1]. More information about how the data is been processed can be found in Christy et al. (2000). Satellites do not measure temperature directly but measure radiances in various wavelength bands and then mathematically inverted to obtain the actual temperature.

[1] `https://www.ncdc.noaa.gov/temp-and-precip/msu/overview`

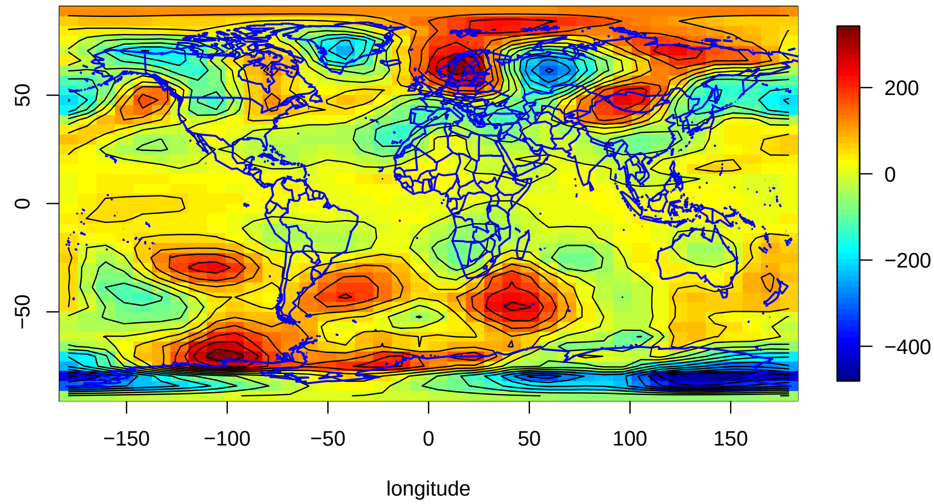

Figure 2.2: MSU data observed (without removing any spatial trends) in August 2002 : resolution 2.5° latitude × 2.5° longitude. It is easy to observe that variation is higher towards the north and south poles.

The MSU data were observed at 2.5°latitude × 2.5°longitude with total number of data observations of size 10368.

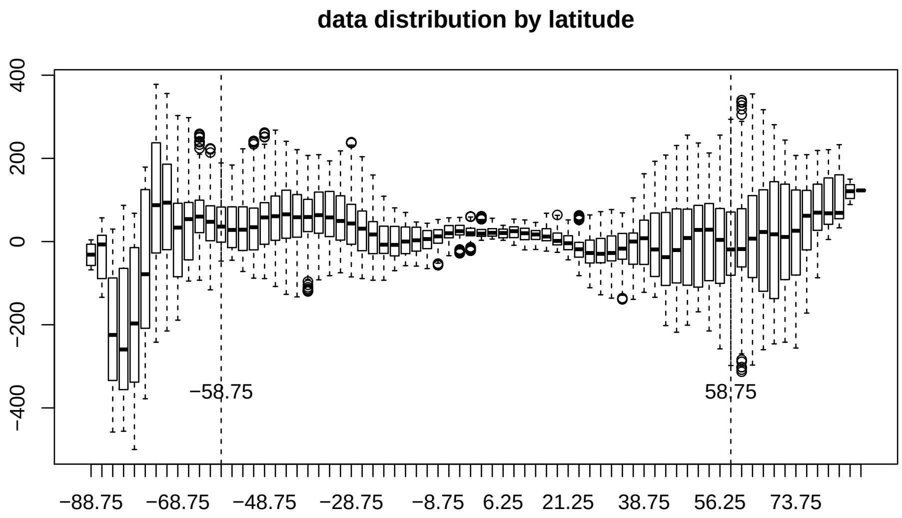

Figure 2.3: MSU data (August 2002), data distribution at each latitude (the spread is very high near north and south poles)

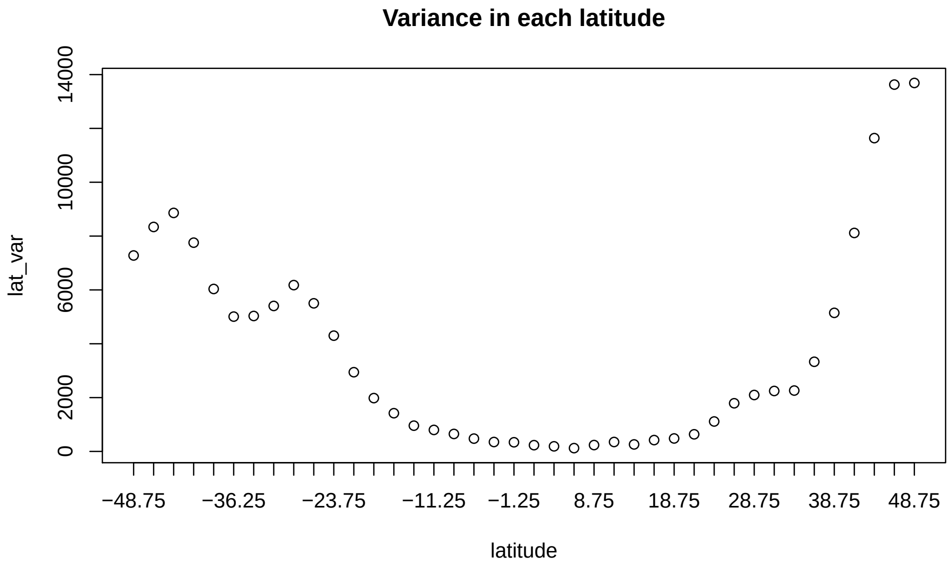

Figure 2.4: MSU data (August 2002) between $60°S$ and $60°N$, the variance at each latitude (variance is almost zero near the equator)

2.1.2 TOMS data

Extracted from Stein (2007) "The Nimbus-7 satellite carried a Total Ozone Mapping Spectrometer (TOMS) instrument that measured total column ozone daily from November 1, 1978 to May 6, 1993. This satellite followed a Sun-synchronous polar orbit with an orbital frequency of 13.825 orbits a day (cycle time about 104 minutes). As the satellite orbited, a scanning mirror repeatedly scanned across a track about 3000 km wide, each track yielding 35 total column ozone measurements. This version of the data is known as Level 2 and is publicly available [2]. "

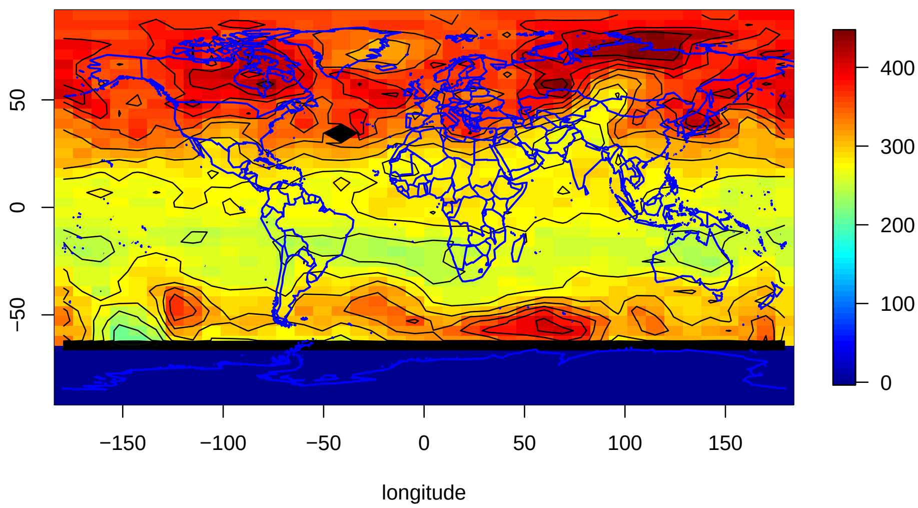

Figure 2.5: TOMS data: resolution 1°latitude × 1.25° longitude in May, 1-6 1990. The instrument used backscattered sunlight, therefore measurements were not available south of 73°S during this week.

There were some missing values in this data set. Stein (2007) used the average of 8 neighboring locations to replace the missing values. Further, he used spherical harmonics with associated Legendre polynomials of up to 78 covariates to remove the spatial trends to study axial symmetry of the data.

[2] `http://disc.sci.gsfc.nasa.gov/data/datapool/TOMS/Level2/`

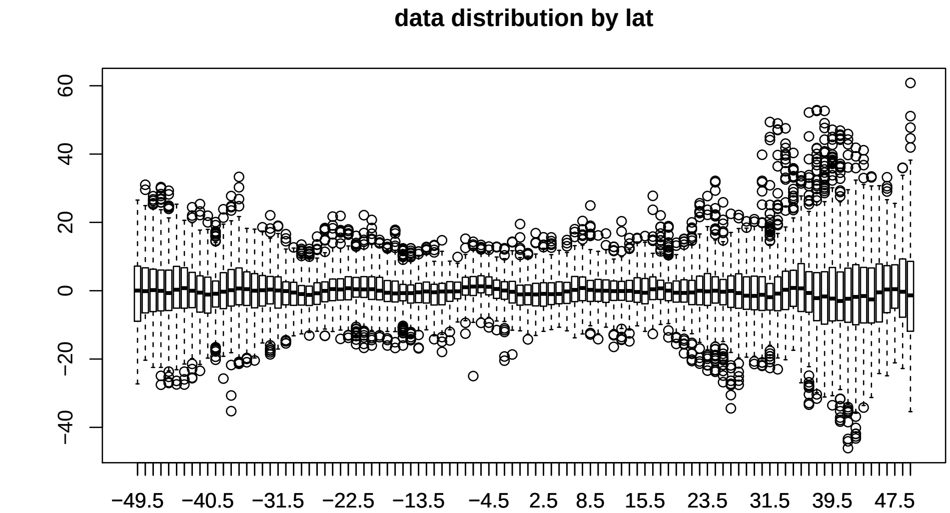

Figure 2.6: TOMS data: data distribution at each latitude (data between 50°S and 50°N were considered)

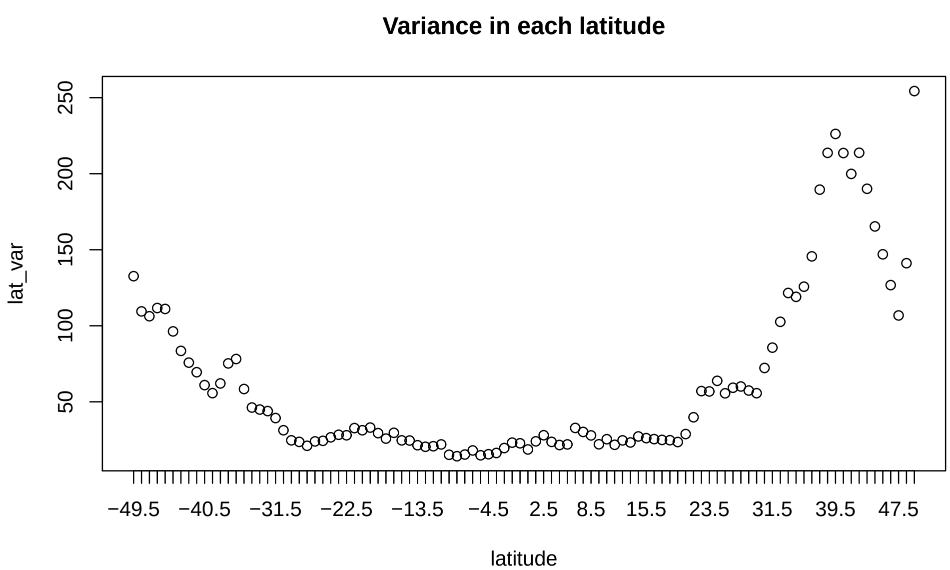

Figure 2.7: TOMS data: between 50°S and 50°N, the variance at each latitude (variance is almost zero near the equator)

2.1.3 Challenges

1. There have been extensive statistical research on methodologies and techniques developed under the Euclidean space R^d. These approaches that are valid in R^d have been

applied to analyze global-scale data in recent years, due to global networks and satellite sensors that have been used to monitor a wide array of global-scale processes and variables. However, this can have unforeseen impacts, such as making use of models that are valid in R^d but in fact might not be valid under spherical coordinate systems. Huang et al. (2011) have investigated some of commonly used covariance models that are valid in R^d, and they pointed out that many of those are actually invalid on the sphere.

2. Note that, as we will see later, the spectral representation of the process on the sphere is a summation of Legendre polynomials, which is distinct from its planar counterpart as represented by an integration of Bessel functions. This distinction can be understood through group representation theory, which is the basis for an exciting new line of research we are currently pursuing.

Organize!!!

1. Axially symmetry, which means that a process is invariant to rotations about the Earth's axis. The idea was first proposed by Jones (1963), where the covariance function depends on the longitudes only through their difference.

2. In the study of a random process on a sphere, homogeneity (covariance depends solely on distance between locations) was assumed. However, this assumption may not be reasonable for actual data. Stein (2007) argued that Total Ozone Mapping Spectrometer (TOMS) data varies strongly with latitudes and homogeneous models are not suitable. Further, Cressie and Johannesson (2008), Jun and Stein (2008), Bolin and Lindgren (2011) pointed out that homogeneity assumption is not reasonable.

3. There are no methods to test axially symmetry in real data. However, this assumption is more plausible and reasonable when modeling spatial data. For example, temperature, moisture, etc. most likely symmetric on longitudes rather than latitudes. Stein (2007) propose a method to model axially symmetric process on a sphere (the fitted model is not the best, but this was a good start).

4. There are no practically useful parametric models available, for our knowledge only models available so far, Stein (1999) with 170 parameters to estimate and Cressie and Johannesson (2008) more than 396 parameters to estimate.

5. When modeling spatial data stationary models are less useful;Jun and Stein (2008) has proposed flexible class of parametric covariance models to capture the non-stationarity of global data. They used Discrete Fourier Transform (DFT) to the data on regular grids and calculated the exact likelihood for large data sets. Furthermore, they used Legendre polynomials to remove the spatial trends when fitting models to global data.

6. Lindgren et al. (2011) analyzed global temperature data with a non-stationary model defined on a sphere using Gaussian Markov Random Fields (GMRF) and Stochastic Partial Differential Equations (SPDE)

7. Monte Carlo Markov Chain (MCMC) is another approach to model non-stationary covariance models on a sphere. Bolin and Lindgren (2011) (continuation of the work proposed in Lindgren et al. (2011)) constructed a class of stochastic field models using Stochastic Partial Differential Equations (SPDEs). Non stationary covariance models were obtained by spatially varying the parameters in the SPDEs, they argue that this method is more efficient than standard MCMC procedures. There are many articles followed this techniques but we will not discuss more details about these methods.

8. Spatio-temporal mixed-effects model for dimension reduction was proposed by Katzfuss and Cressie (2011). They used MOM parameter estimation method (similar approach in FRS). This work is also based on Cressie and Johannesson (2008) spatial only Fixed Rank model. These methods are eventually focused on Bayesian approach and are less interested about topic.

9. The previous studies have argued that many processes on a sphere are not homogeneous, especially in latitude direction. Huang et al. (2012) proposed a class of statistical processes that are axially symmetric and covariance functions that depend on longitudinal differences. Moreover, they have proposed longitudinally reversible processes and some motivations to construct axially symmetric processes. The covariance models implemented in this dissertation are modified versions of the covariance models proposed by Huang et al. (2012).

10. Hitczenko and Stein (2012) discuss about the properties of an existing class of models for axially symmetric Gaussian processes on the sphere. They applied first-order differential operators to an isotropic process. draw conclusions about the local properties of the processes. Under some restrictions they derived explicit forms for the spherical harmonic representation of these processes covariance functions, and make conclusions about the local properties of the processes.

11. The issues associated when modeling axially symmetric spatial random fields on a sphere was discussed by Li (2013). They proposed convolution methods to generate random fields with a class of $Mat\acute{e}rn$-type kernel functions by allowing the parameters in the kernel function to vary with location. Moreover, they were able to generate flexible class of covariance functions and capture the non-stationary properties on a sphere. Used FFT to get the determinant and the inverse efficiently. Further, semi-parametric variogram estimation method using spectral representation was proposed for intrinsically stationary random fields on S^2.

12. $Mat\acute{e}rn$ covariance models are widely used when modeling spatial data, but when the smoothness parameter (ν) is greater than 0.5 it is not valid for the homogeneous processes on the Earth surface with great circle distance. Jeong and Jun (2015) proposed $Mat\acute{e}rn$-like covariance functions for smooth processes on the earth surface that are valid with great circle distance (models were tested on sea levels pressure data).

Family	C(h)	Parameters	Validity
$Matérn$	$\frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\frac{h}{\phi})^{\nu}Y_{\nu}(\frac{h}{\phi})$	ν, σ^2, ϕ	R^3, S^2 when $\nu \leq 0.5$
Spherical	$\sigma^2(1 - \frac{3h}{2\phi} + \frac{1}{2}(\frac{h}{\phi})^3)I_{0\leq h\leq\phi}$	ϕ, σ^2	R^3, S^3
Exponential	$\sigma^2 exp\{-(h/\phi)\}$	ϕ, σ^2	R^3
Gaussian	$\sigma^2 exp\{-(h/\phi)^2\}$	ϕ, σ^2	R^3
Power	$\sigma^2(C_0 - (h/\phi)^{\alpha}$	ϕ, σ^2	$R^3\alpha \in [0,2], S^2\alpha \in (1,2]$

Table 2.1: Commonly used covariance and variogram models

Chapter 3

Covariance and Variogram Estimation on the Cricle

1. Introduction. Time series and spatial statistics with Noel Cressie's results

3.1 Random process on a circle

Let $X(t)$ be the random process on the unit circle S^1. If $X(t)$ is further assumed to be with finite second moment and continuity in quadratic mean, then $X(t)$ can be represented in a Fourier series which is convergent in quadratic mean (Dufour and Roy (1976)),

$$X(t) = A_0 + \sum_{n=1}^{\infty}(A_n \cos(nt) + B_n \sin(nt)), \qquad t \in S^1,$$

where

$$A_0 = \frac{1}{2\pi}\int_S X(t)dt, \quad A_n = \frac{1}{\pi}\int_S X(t)\cos(nt)dt, \quad B_n = \frac{1}{\pi}\int_S X(t)\sin(nt)dt.$$

Let $s, t \in S^1$. The covariance function $C(s,t)$ of the process $X(t)$ on the given locations s and t is given below

$$C(s,t) = cov(X(s), X(t)).$$

Now we assume the underlying process $X(t)$ is stationary on the circle, that is, $E(X(t)) = \mu$ unknown, and its covariance function solely depends on the angular distance θ

$$C(\theta) = cov(X(t+\theta), X(t)), \qquad \theta \in [0, \pi].$$

Under the assumption of stationarity, we have

$$cov(A_n, A_m) = cov(B_n, B_m) = a_n\delta(n,m), \qquad \text{and} \qquad cov(A_n, B_m) = 0, \quad \text{for } n \geq 0, m > 0,$$

with $a_n \geq 0$, under which the covariance function $C(\theta)$ can be written as the following spectral representation

$$C(\theta) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\theta), \qquad \text{for } \theta \in [0, \pi].$$

By the orthogonality of $\{\cos(n\theta), n = 0, 1, 2, \cdots,\}$ on $\theta \in [0, \pi]$, we have

$$a_0 = \frac{1}{\pi} \int_0^{\pi} C(\theta) d\theta, \qquad a_n = \frac{2}{\pi} \int_0^{\pi} C(\theta) \cos(n\theta) d\theta, \quad n \geq 1.$$

If a random process $X(t)$ is intrinsically stationary, one has $E(X(t)) = \mu$, an unknown constant, as well as the variogram function depends only on the angular distance θ, given by

$$\gamma(\theta) = var(X(t + \theta) - X(t)), \qquad t \in S^1.$$

Note that if $X(t)$ is stationary, then

$$\gamma(\theta) = C(0) - C(\theta).$$

Equivalently, $\gamma(\theta)$ has the following spectral representation

$$\gamma(\theta) = \sum_{n=1}^{\infty} a_n (1 - \cos(n\theta)).$$

3.1.1 Mean and covariance estimation on the circle

The ultimate goal in the spatial statistics is to make prediction of random process on the unobserved location. We now consider the estimation on the unknown mean μ and covariance function $C(\theta)$. Let $\{X(t_k), k = 1, 2, \cdots, n\}$ be a collection of gridded observations on the unit circle, with $t_k = (k - 1) * 2\pi/n, k = 1, 2, \cdots, n$. For simplicity, let $n = 2N$ be an even number. Denote $\underset{\sim}{X} = (X_1, X_2, \cdots, X_n)^T$ as the observed random vector. When the underlying process $X(t)$ is stationary on the unit circle, the variance-covariance matrix of $\underset{\sim}{X}$ is given by

$$\Sigma = \begin{pmatrix} C(0) & \cdots & C((N-1)\delta) & C(\pi) & C((N-1)\delta) & \cdots & C(\delta) \\ C(\delta) & \cdots & C((N-2)\delta) & C((N-1)\delta) & C(\pi) & \cdots & C(2\delta) \\ C(2\delta) & \cdots & C((N-3)\delta) & C((N-2)\delta) & C((N-1)\delta) & \cdots & C(3\delta) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C(\delta) & \cdots & C(\pi) & C((N-1)\delta) & C((N-2)\delta) & \cdots & C(0) \end{pmatrix},$$

is a symmetric circulant matrix with elements $C(0), C(\delta), C(2\delta), \cdots, C((N-1)\delta), C(\pi), C((N-1)\delta), \cdots, C(\delta)$ where $\delta = 2\pi/n$. Therefore the sample mean

$$\bar{X} = \frac{1}{n} \mathbf{1}_n^T \underset{\sim}{X}$$

is an unbiased estimator of μ with the variance given by

$$\begin{aligned} var(\bar{X}) &= cov(\frac{1}{n}\mathbf{1}_n^T \underset{\sim}{X}, \frac{1}{n}\mathbf{1}_n^T \underset{\sim}{X}) \\ &= \frac{1}{n^2}\mathbf{1}_n^T \Sigma \mathbf{1}_n \\ &= \frac{1}{n}\left(C(0) + C(\pi) + 2\sum_{m=1}^{N-1} C(m2\pi/n) \right) \end{aligned}$$

If we assume that $C(\theta)$ is a continuous function on $[0, \pi]$ and note the summation in the last quantity is a trapezoid sum of $C(\theta)$ on the gridded locations within $[0, \pi]$, then,

$$\frac{1}{\pi}\frac{\pi}{2N}\left(C(0) + \sum_{m=1}^{N-1} C(m2\pi/n) + C(\pi) \right) \to \frac{1}{\pi}\int_0^{\pi} C(\theta) d\theta = a_0,$$

as $n \to \infty$. That is, $var(\bar{X}) \to a_0$ as $n \to \infty$. Therefore, we have the following proposition.

Proposition 3.1.1 *The sample mean $\bar{X}$ is an unbiased estimator of μ with the asymptotic variance of a_0. If $a_0 > 0$ and $X(t)$ is further assumed to be Gaussian, then $\bar{X}$ is NOT a consistent estimator of μ.*

Proof: *It is only necessary to prove the second part. If $X(t)$ is Gaussian, then $\bar{X} \sim N(\mu, var(\bar{X})) \Rightarrow Z = \frac{\bar{X}-\mu}{\sqrt{var(\bar{X})}} \sim N(0,1)$. First, for a fixed $\varepsilon_0 > 0$ and $\varepsilon_0 < a_0$, there exists K, such that for all $n > K$, we have*

$$|var(\bar{X}) - a_0| < \varepsilon_0 \Rightarrow a_0 - \varepsilon_0 < var(\bar{X}) < a_0 + \varepsilon_0.$$

Now for each fixed $\varepsilon > 0$ and all $n > K$,

$$\begin{aligned} P\left(|\bar{X} - \mu| > \varepsilon\right) &= P\left(\frac{|\bar{X} - \mu|}{\sqrt{var(\bar{X})}} > \frac{\varepsilon}{\sqrt{var(\bar{X})}} \right) \\ &\geq P\left(|Z| > \frac{\varepsilon}{\sqrt{a_0 - \varepsilon_0}} \right) > 0. \end{aligned}$$

Hence $\bar{X} \not\to \mu$ in probability. The last inequality above is due to the following.

$$\left\{ |Z| > \frac{\varepsilon}{\sqrt{a_0 - \varepsilon_0}} \right\} \subseteq \left\{ |Z| > \frac{\varepsilon}{\sqrt{var(\bar{X})}} \right\}.$$

Now we consider the MOM estimator of $C(\theta)$, given by

$$\hat{C}(\Delta\lambda) = \frac{1}{n}\sum_{i=1}^{n}(X(t_i + \Delta\lambda) - \bar{X})(X(t_i) - \bar{X}), \tag{3.1.1}$$

where $\Delta\lambda = 0, 2\pi/n, 4\pi/n, \cdots, 2(N-1)\pi/n$.

Now we calculate the unbiasedness and consistency of the above estimator.

$$\begin{aligned}
E(\hat{C}(\Delta\lambda)) &= \frac{1}{n}\sum_{i=1}^{n} E((X(t_i+\Delta\lambda)-\bar{X})(X(t_i)-\bar{X})) \\
&= \frac{1}{n}\sum_{i=1}^{n} E((X(t_i+\Delta\lambda)-\mu-(\bar{X}-\mu))(X(t_i)-\mu-(\bar{X})-\mu)) \\
&= \frac{1}{n}\sum_{i=1}^{n} cov(X(t_i+\Delta\lambda), X(t_i)) - \frac{1}{n}\sum_{i=1}^{n} E((X(t_i+\Delta\lambda)-\mu)(\bar{X}-\mu)) \\
&\quad -\frac{1}{n}\sum_{i=1}^{n} E((X(t_i)-\mu)(\bar{X}-\mu)) + \frac{1}{n}\sum_{i=1}^{n} E((\bar{X}-\mu)(\bar{X}-\mu)) \\
&= C(\Delta\lambda) - E((\bar{X}-\mu)(\bar{X}-\mu)) - E((\bar{X}-\mu)(\bar{X}-\mu)) \\
&\quad +E((\bar{X}-\mu)(\bar{X}-\mu)) \\
&= C(\Delta\lambda) - var(\bar{X}).
\end{aligned}$$

That is, the MOM estimator $\hat{C}(\Delta\lambda)$ of the covariance function is actually a biased estimator with the shift amount approximately equal to a_0 unless $a_0 = 0$. In other words, if $a_0 = 0$, the MOM estimator $\hat{C}(\Delta\lambda)$ is an unbiased estimator of $C(\theta)$. In summary, we have

Proposition 3.1.2 *The MOM covariance estimator is a biased estimator of the true covariance function $C(\theta)$, if $a_0 > 0$. However, if the process is a zero mean process or $\mu = 0$ then the MOM covariance estimator is unibiased.*

Remark 1: $a_0 = 0$ implies that $var(\bar{X}) \to 0 \Rightarrow \bar{X} \to \mu$ a.s.. On the other hand, if $a_0 \neq 0$, then $var(\bar{X}) \neq 0$. Proposition 3.1.1 indicates that $\bar{X}$ will never be a consistent estimator for μ. Note that in the time series case when $X(t), t = 0, \pm1, \pm2, \cdots$ is stationary, $E(\bar{X}-\mu)^2 \to 0$ as $n \to \infty$ under the ergodic assumption that the covariance function $C(t) \to 0$ when $t \to \infty$ (which is practically feasible), that is, $\bar{X}$ is consistent in the time series case. Note that on the case of circle, in particular, in the case of exponential covariance function $C(\theta) = C_1 e^{-a\theta}, a > 0, C_1 > 0$,

$$a_0 = \frac{C_1}{a\pi}(1 - e^{-a\pi})$$

and so $a_0 > 0$ all the time. Generally in the case of circle, we may not have $C(\theta)$ close to 0 since θ is within a bounded region ($[0, \pi]$ for the circle) and we normally assume $C(\theta)$ is continuous for θ.

Remark 2: If in practice, we have multiple copies of data observations on the circle, we can then estimate a_0 or $var(\bar{X})$ through these copies.

Suppose we have *i.i.d.* copies of the same data on the circle saying that the averages are $\{\bar{X}_i, i = 1, 2, \cdots, m\}$. We consider the estimator for a_0. Obviously, we use the MOM given as following:

$$\hat{a}_0 = \frac{1}{m-1}\sum_{j=1}^{m}(\bar{X}_j - \bar{\bar{X}})^2,$$

where $\bar{\bar{X}} = \frac{1}{m}\sum_{k=1}^{m}\bar{X}_k$. Given m *i.i.d.* copies of data $Y_1, Y_2, \cdots, Y_m$ (Here $Y_i = \bar{X}_i$ in the above setting), $E(Y_i) = \mu$ unknown, and $var(Y_i) = a_0$ is also unknown. Then the general estimation method to estimate a_0 is the sample variance

$$S^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2$$

with $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$. Obviously, S^2 is an unbiased estimator of $var(Y_i) = a_0$.

$$var(S^2) = O(1/m) \to 0 \quad \text{as } m \to \infty.$$

Here we need to assume the fourth moment $E(Y_i^4)$ exists. Hence, S^2 is also consistent.

Further under the assumption of normality, we have

$$\frac{(m-1)S^2}{a_0} \sim \chi^2_{m-1}$$

Hence

$$\begin{aligned} var\left(\frac{(m-1)S^2}{a_0}\right) &= 2(m-1) \\ var(S^2) &= \frac{a_0^2}{m-1} \to 0 \quad \text{as } m \to \infty. \end{aligned}$$

This also shows the consistency of the estimator S^2. Back to the circle case, the normality assumption is reasonable since we use multivariate normal random variates when we generate data on the circle. $\bar{X}$ is a linear combination of multivariate normal random variables, and hence $\bar{X}$ is normally distributed.

3.1.2 Variogram estimation

In R^n, The variogram estimator generally performs better than the covariance function estimator (Cressie (1993)). Given gridded data observations $\underset{\sim}{X}$, the variogram estimator through Method of Moments is given by

$$\hat{\gamma}(\Delta\lambda) = \frac{1}{2n}\sum_{i=1}^{n}(X(t_i + \Delta\lambda) - X(t_i))^2, \tag{3.1.2}$$

Then

$$\begin{aligned}
E(\hat{\gamma}(\Delta\lambda)) &= \frac{1}{2n}\sum_{i=1}^{n} E(X(t_i+\Delta\lambda)-X(t_i))^2 \\
&= \frac{1}{2n}\sum_{i=1}^{n} E((X(t_i+\Delta\lambda)-\mu)-(X(t_i)-\mu))^2 \\
&= \frac{1}{2n}\sum_{i=1}^{n} cov(X(t_i+\Delta\lambda)-X(t_i), X(t_i+\Delta\lambda)-X(t_i)) \\
&= \frac{1}{2n}\sum_{i=1}^{n} (cov(X(t_i+\Delta\lambda), X(t_i+\Delta\lambda)) + cov(X(t_i), X(t_i)) - 2cov(X(t_i+\Delta\lambda), X(t_i))) \\
&= \frac{1}{2n}\sum_{i=1}^{n} (C(0)+C(0)-2C(\Delta\lambda)) \\
&= C(0)-C(\Delta\lambda) = \gamma(\Delta\lambda).
\end{aligned}$$

Therefore, $\hat{\gamma}(\Delta\lambda)$ is an unbiased estimator of $\gamma(\theta)$.

We first calculate the variance and covariance of the variogram estimator on the circle. Again we consider the equal-distance gridded points on the circle $\{t_i : 1 \le i \le n, t_i = (i-1)\times 2\pi/n\}$ and $\underset{\sim}{X} = (X(t_1), X(t_2), \ldots, X(t_n))^T$, being the observed data values. Assume that the random process $X(t)$ is stationary. Note that the Matheron's classical semi-variogram estimator on the circle based on the method of moments can be written as

$$\hat{\gamma}(\theta) = \underset{\sim}{X}^T A(\theta) \underset{\sim}{X}.$$

Here for all θ, $A(\theta)$ is a circulant matrix, and in particular, $A(0) = 0$. For simplicity, we set n to be even so that $n = 2N$. Here we give some examples to demonstrate the structure of $A(\theta)$.

Let $n = 6$. We have four distance angles $\theta = 0, \pi/3, 2\pi/3, \pi$. Then each of design matrices

$A(\theta)$ is given below.

$$
\begin{aligned}
A(0) &= 0 \\
A(\pi/3) &= \frac{1}{6}\begin{pmatrix} 2 & -1 & 0 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & 0 & -1 & 2 \end{pmatrix} = \frac{1}{6}circ(2,-1,0,0,0,-1); \\
A(2\pi/3) &= \frac{1}{6}\begin{pmatrix} 2 & 0 & -1 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 & 0 & -1 \\ -1 & 0 & 2 & 0 & -1 & 0 \\ 0 & -1 & 0 & 2 & 0 & -1 \\ -1 & 0 & -1 & 0 & 2 & 0 \\ 0 & -1 & 0 & -1 & 0 & 2 \end{pmatrix} = \frac{1}{6}circ(2,0,-1,0,-1,0); \\
A(\pi) &= \frac{1}{6}\begin{pmatrix} 2 & 0 & 0 & -2 & 0 & 0 \\ 0 & 2 & 0 & 0 & -2 & 0 \\ 0 & 0 & 2 & 0 & 0 & -2 \\ -2 & 0 & 0 & 2 & 0 & 0 \\ 0 & -2 & 0 & 0 & 2 & 0 \\ 0 & 0 & -2 & 0 & 0 & 2 \end{pmatrix} = \frac{1}{6}circ(2,0,0,-2,0,0);
\end{aligned}
$$

Note that for a circulant matrix $M = circ(c_0, c_1, c_2, \ldots, c_{n-2}, c_{n-1})$, the eigenvalues are given by

$$\lambda_j = c_0 + c_1\omega_j + c_2\omega_j^2 + \cdots + c_{n-1}\omega_j^{n-1}, \quad j = 0, 1, 2, \cdots, n-1,$$

and $\omega_j = \exp(j2\pi i/n), i^2 = -1$.

In addition, all circulant matrices can be orthogonally diagonalized using the same orthogonal matrix (the so-called Fourier matrix). Hence, the trace of the product of circulant matrices = the trace of product of diagonal matrices = the sum of the product of corresponding eigenvalues from those circulant matrices.

Now we consider the distribution of the variogram estimator. First we write the variogram estimator in the following form

$$
\begin{aligned}
\hat{\gamma}(\Delta\lambda) &= \frac{1}{2n}\sum_{i=1}^{n}(X(t_i + \Delta\lambda) - X(t_i))^2 \\
&= \frac{1}{2n}\sum_{i=1}^{n}((X(t_i + \Delta\lambda) - \mu) - (X(t_i) - \mu))^2.
\end{aligned}
$$

Therefore,

$$\hat{\gamma}(\Delta\lambda) = (\utilde{X} - \utilde{\mathbf{1}}_n\mu)^T A(\Delta\lambda)(\utilde{X} - \utilde{\mathbf{1}}_n\mu) \tag{3.1.3}$$

Note that $A(\Delta\lambda)$ is a circulant matrix with following spectral decomposition

$$A(\Delta\lambda) \quad = \quad P\Lambda^{(A)}P^T,$$

where P is the so-called fourier matrix (orthogonal), solely depending on the dimension of A, and

$$\Lambda^{(A)} = \text{diag}(\lambda_1^{(A)}, \lambda_2^{(A)}, \cdots, \lambda_n^{(A)})$$

with

$$\lambda_m^{(A)} = \frac{1}{n}(1 - \cos((m-1)\Delta\lambda)), \qquad m = 1, 2, \cdots, n.$$

If $\underset{\sim}{X}$ follows a multivariate normal $N(\underset{\sim}{\mathbf{1}}_n\mu, \Sigma)$, then $(\underset{\sim}{X} - \underset{\sim}{\mathbf{1}}_n\mu) \sim N(\underset{\sim}{\mathbf{0}}, \Sigma)$. Note that the variance-covariance matrix Σ is also a circulant matrix, which has the following spectral decomposition.

$$\Sigma = P\Lambda^{(\Sigma)}P^T,$$

$$\text{with } \Lambda^{(\Sigma)} = \text{diag}(\lambda_1^{(\Sigma)}, \lambda_2^{(\Sigma)}, \cdots, \lambda_n^{(\Sigma)}),$$

$$\text{where } \lambda_j^{(\Sigma)} = \left(C(0) + 2\sum_{m=1}^{N-1} C(m\delta)\cos((j-1)m\delta) + C(\pi)\cos((j-1)N\delta) \right).$$

Let $\underset{\sim}{Y} = P^T\left(\underset{\sim}{X} - \underset{\sim}{\mathbf{1}}_n\mu\right)$, then $\underset{\sim}{Y}$ follows a multivariate normal distribution with mean $\underset{\sim}{\mathbf{0}}$ and variance-covariance matrix given by

$$\begin{aligned} var(\underset{\sim}{Y}) &= cov(P^T\left(\underset{\sim}{X} - \underset{\sim}{\mathbf{1}}_n\mu\right), P^T\left(\underset{\sim}{X} - \underset{\sim}{\mathbf{1}}_n\mu\right)) \\ &= P^T\Sigma P = P^T P\Lambda^{(\Sigma)}P^T P = \Lambda^{(\Sigma)}. \end{aligned}$$

That is, $\underset{\sim}{Y} = (Y_1, Y_2, \cdots, Y_n)^T$ are independent normal random variates with mean 0 and variance $\lambda_j^{(\Sigma)}$.

The variogram estimator is then given by

$$\begin{aligned} \hat{\gamma}(\Delta\lambda) &= (\underset{\sim}{X} - \underset{\sim}{\mathbf{1}}_n\mu)^T A(\Delta\lambda)(\underset{\sim}{X} - \underset{\sim}{\mathbf{1}}_n\mu) \\ &= (P(\underset{\sim}{X} - \underset{\sim}{\mathbf{1}}_n\mu))^T\Lambda^{(A)}(P^T(\underset{\sim}{X} - \underset{\sim}{\mathbf{1}}_n\mu)) \\ &= \underset{\sim}{Y}\Lambda^{(A)}\underset{\sim}{Y} = \sum_{m=1}^{n}\lambda_m^{(A)}Y_m^2. \end{aligned}$$

Note $\frac{Y_m}{\sqrt{\lambda_m^{(\Sigma)}}} \sim N(0,1)$, and so $\frac{Y_m^2}{\lambda_m^{(\Sigma)}} \sim \chi_1^2$ (or written as $\chi_{1,m}^2$ due to the dependency on m), which implies

$$\hat{\gamma}(\Delta\lambda) \quad = \quad \sum_{m=1}^{n}\lambda_m^{(A)}\lambda_m^{(\Sigma)}\left(\frac{Y_m}{\sqrt{\lambda_m^{(\Sigma)}}}\right)^2 \sim \sum_{m=1}^{n}\lambda_m^{(A)}\lambda_m^{(\Sigma)}\chi_{1,m}^2.$$

Here $\chi^2_{1,1}, \chi^2_{1,2}, \cdots, \chi^2_{1,n}$ are *i.i.d.* following χ^2_1 distribution. Hence

$$E(\hat{\gamma}(\Delta\lambda)) = \sum_{m=1}^{n} \lambda_m^{(A)} \lambda_m^{(\Sigma)}, \qquad var(\hat{\gamma}(\Delta\lambda)) = 2\sum_{m=1}^{n} (\lambda_m^{(A)} \lambda_m^{(\Sigma)})^2$$

We showed that the mean of the variogram estimator is unbiased. Now we will show the variance of the variogram estimator does not converge to zero for each $\Delta\lambda$.

First notice the unbiasedness of the variogram estimator

$$E(\hat{\gamma}(\Delta\lambda)) = \gamma(\Delta\lambda) = C(0) - C(\Delta\lambda),$$

that is,

$$\sum_{m=1}^{n} \lambda_m^{(A)} \lambda_m^{(\Sigma)} = C(0) - C(\Delta\lambda).$$

Hence

$$\begin{aligned} \hat{\gamma}(\Delta\lambda) &= \sum_{m=1}^{n} \lambda_m^{(A)} \lambda_m^{(\Sigma)} \chi^2_{1,m} \\ &= (C(0) - C(\Delta\lambda)) \sum_{m=1}^{n} \frac{\lambda_m^{(A)} \lambda_m^{(\Sigma)}}{C(0) - C(\Delta\lambda)} \chi^2_{1,m} \\ &\triangleq (C(0) - C(\Delta\lambda)) \sum_{m=1}^{n} C_{n,m} \chi^2_{1,m}, \end{aligned}$$

where $\sum_{m=1}^{n} C_{n,m} = \sum_{m=1}^{n} \frac{\lambda_m^{(A)} \lambda_m^{(\Sigma)}}{C(0)-C(\Delta\lambda)} = 1$ and $C_{n,m} > 0$ since both the matrices A and Σ are positive definite. Hence

$$\begin{aligned} var(\hat{\gamma}(\Delta\lambda)) &= (C(0) - C(\Delta\lambda))^2 * 2 * \left(\sum_{m=1}^{n} C_{n,m}^2 \right) \\ &\leq 2(C(0) - C(\Delta\lambda))^2 \left(\sum_{m=1}^{n} C_{n,m} \right) = 2(C(0) - C(\Delta\lambda))^2. \end{aligned}$$

On the other hand,

$$
\begin{aligned}
var(\hat{\gamma}(\Delta\lambda)) &= (C(0) - C(\Delta\lambda))^2 * 2 * \left(\sum_{m=1}^{n} C_{n,m}^2 \right) \\
&\geq (C(0) - C(\Delta\lambda))^2 * 2 * C_{n,2}^2 \\
&= 2\frac{1}{n^2}(1 - \cos(\Delta\lambda))^2 \left(C(0) + 2\sum_{k=1}^{N-1} C(k\delta)\cos(k\delta) + C(\pi)\cos(N\delta) \right)^2 \\
&\sim 2(1 - \cos(\Delta\lambda))^2 \frac{4\pi^2}{n^2} \left(\frac{2}{4\pi^2} \sum_{k=0}^{N} C(k\delta)\cos(k\delta) \right)^2 \\
&\sim (1 - \cos(\Delta\lambda))^2 \left(\frac{1}{\pi} \int_0^{\pi} C(\theta)\cos(\theta)d\theta \right)^2 \\
&= \frac{a_1^2}{4}(1 - \cos(\Delta\lambda))^2.
\end{aligned}
$$

In summary,

$$
\frac{a_1^2}{4}(1 - \cos(\Delta\lambda))^2 \leq var(\hat{\gamma}(\Delta\lambda)) \leq 2(C(0) - C(\Delta\lambda))^2.
$$

We consider the consistency of the variogram estimator, that is, we consider the following limit

$$
P\left(|\hat{\gamma}(\Delta\lambda) - \gamma(\Delta\lambda)| \geq \varepsilon\right) \to 0,
$$

as $n \to \infty$ for fixed $\varepsilon > 0$ and $\Delta\lambda \neq 0$ (note here $\Delta\lambda = k * 2\pi/n$ and $k/n \to$ a constant as $n \to \infty$ to maintain that we are estimating a fixed distance $\theta = \Delta\lambda$). Notice that

$$
\hat{\gamma}(\Delta\lambda) = (C(0) - C(\Delta\lambda)) \sum_{m=1}^{n} C_{n,m}\chi_{1,m}^2, \quad \text{and} \quad \gamma(\Delta\lambda) = C(0) - C(\Delta\lambda),
$$

hence,

$$
\hat{\gamma}(\Delta\lambda) - \gamma(\Delta\lambda) \;=\; (C(0) - C(\Delta\lambda)) \left(\sum_{m=1}^{n} C_{n,m}\chi_{1,m}^2 - 1 \right),
$$

with $\sum_{m=1}^{n} C_{n,m} = 1$ and $C_{n,m} > 0$ for any n. Therefore, if $\Delta\lambda \neq 0$, the consistency of $\hat{\gamma}(\Delta\lambda)$ is equivalent to the consistency of the random sum $\sum_{m=1}^{n} C_{n,m}\chi_{1,m}^2$ converging to 1. In literature, it seems that there is no closed expression for the linear combination (or weighted) chi-square random variables.

As some examples, we exam

$$
\begin{aligned}
C_{n,1} &= 0 \\
C_{n,2} &= \frac{1}{n}(1-\cos(\Delta\lambda))\left(C(0)+2\sum_{k=1}^{N-1}C(k\delta)\cos(k\delta)+C(\pi)\cos(\pi)\right)/(C(0)-C(\Delta\lambda)) \\
&\to (1-\cos(\Delta\lambda))\left(\frac{1}{\pi}\int_0^{\pi}C(\theta)\cos(\theta)d\theta\right)/(C(0)-C(\Delta\lambda)) \\
&= \frac{a_1}{2}(1-\cos(\Delta\lambda))/(C(0)-C(\Delta\lambda)) \\
C_{n,5} &= \frac{1}{n}(1-\cos(4\Delta\lambda))\left(C(0)+2\sum_{k=1}^{N-1}C(k\delta)\cos(4k\delta)+C(\pi)\cos(4\pi)\right)/(C(0)-C(\Delta\lambda)) \\
&\to (1-\cos(4\Delta\lambda))\left(\frac{1}{\pi}\int_0^{\pi}C(\theta)\cos(4\theta)d\theta\right)/(C(0)-C(\Delta\lambda)) \\
&= \frac{a_4}{2}(1-\cos(4\Delta\lambda))/(C(0)-C(\Delta\lambda)).
\end{aligned}
$$

In general, for a fixed m, we have

$$
\begin{aligned}
C_{n,m} &= \frac{1}{n}(1-\cos((m-1)\Delta\lambda))\left(C(0)+2\sum_{k=1}^{N-1}C(k\delta)\cos((m-1)k\delta)+C(\pi)\cos((m-1)\pi)\right) \\
&\quad /(C(0)-C(\Delta\lambda)) \\
&\to (1-\cos((m-1)\Delta\lambda))\left(\frac{1}{\pi}\int_0^{\pi}C(\theta)\cos((m-1)\theta)d\theta\right)/(C(0)-C(\Delta\lambda)) \\
&= \frac{a_{m-1}}{2}(1-\cos((m-1)\Delta\lambda))/(C(0)-C(\Delta\lambda)), \quad \text{as } n\to\infty.
\end{aligned}
$$

Proposition 3.1.3 *The MOM estimator of variogram function on the circle is unbiased with bounded away from zero variance. In addition, if the underlying process $X(t)$ is Gaussian, the variogram estimator is not consistent on the circle.*

Proof: *First we consider the consistency of the variogram estimator. To show the following*

$$P\left(|\hat{\gamma}(\Delta\lambda)-\gamma(\Delta\lambda)|\geq\varepsilon\right)\to 0,$$

as $n\to\infty$ for fixed $\varepsilon>0$ and $\Delta\lambda\neq 0$, it is equivalent to show that

$$P\left(\left|\sum_{m=1}^{n}C_{n,m}\chi^2_{1,m}-1\right|\geq\varepsilon\right)\to 0,$$

as $n\to\infty$ for fixed $\varepsilon>0$ and $\Delta\lambda\neq 0$. Here $\sum_{m=1}^{n}C_{n,m}=1, C_{n,m}>0$ for each fixed n. Note that we also have for each fixed m,

$$0<C_{n,m}\to\frac{a_m}{2}\frac{1-\cos(m\Delta\lambda)}{C(0)-C(\Delta\lambda)}\equiv b_m.$$

For simplicity, we can assume that $b_2 > 0$ (Otherwise we can pick some $b_m > 0$ for some m fixed). That is

$$C_{n,2} \to b_2 > 0, \qquad \text{as } n \to \infty.$$

Therefore, for fixed $\varepsilon_0 > 0$ and $\varepsilon_0 < b_2$, we choose all $n > N$, such that

$$b_2 - \varepsilon_0 < C_{n,2} < b_2 + \varepsilon_0$$

Therefore, for all $n > N$, (and denote $\chi^2_{1,2} = \chi^2_1$ for simplicity)

$$\sum_{m=1}^{n} C_{n,m}\chi^2_{1,m} \geq C_{n,2}\chi^2_{1,2} > (b_2 - \varepsilon_0)\chi^2_1$$

Hence notice that, for the fixed $\varepsilon > 0$,

$$\left\{(b_2 - \varepsilon_0)\chi^2_1 > 1 + \varepsilon\right\} \subseteq \left\{\sum_{m=1}^{n} C_{n,m}\chi^2_{1,m} > 1 + \varepsilon\right\}$$

Now, for all $n \geq N$,

$$\begin{aligned}
& P\left(\left|\sum_{m=1}^{n} C_{n,m}\chi^2_{1,m} - 1\right| \geq \varepsilon\right) \\
= \; & P\left(\sum_{m=1}^{n} C_{n,m}\chi^2_{1,m} > 1 + \varepsilon \quad or \quad \sum_{m=1}^{n} C_{n,m}\chi^2_{1,m} < 1 - \varepsilon\right) \\
\geq \; & P\left(\sum_{m=1}^{n} C_{n,m}\chi^2_{1,m} > 1 + \varepsilon\right) \geq P\left((b_2 - \varepsilon_0)\chi^2_1 > 1 + \varepsilon\right) \\
= \; & P\left(\chi^2_1 > \frac{1 + \varepsilon}{b_2 - \varepsilon_0}\right) \not\to 0,
\end{aligned}$$

since the last term is a fixed positive number. This proves the non-consistency of variogram estimator.

3.2 Data generation on a circle

First, we will discuss how to generate correlated data at n (assume even) girdded points on a circle when the covarince function is known and we compare the above covariance and variogram estimators to its theoretical values. Under the assumption of isotropy the covariance function can be written as a function of distance (angle). For the data generation process we consider two covariance fucntions that are valid on a circle, exponential family and power family as given below,

$$C(\theta) = C_1 e^{-a|\theta|} \quad a > 0, C_1 > 0 \tag{3.2.1}$$

$$C(\theta) = c_0 - (|\theta|/a)^{\alpha} \quad a > 0, \alpha \in (0, 2] \text{ and } c_0 \geq \int_0^{\pi} (\theta/a)^{\alpha} \sin\theta d\theta \tag{3.2.2}$$

where $\theta = i * \Delta\lambda = \pm i * 2\pi/n, i = 1, 2, \cdots, \lfloor n/2 \rfloor$

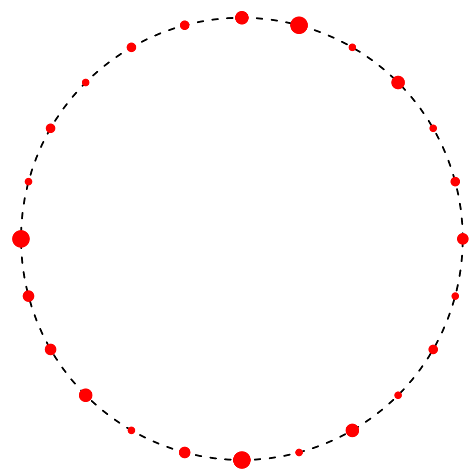

Figure 3.1: Random process on a circle at 24 points ($\Delta\lambda = 15^0$), the red dots represent the observed values at a given time and each point is associated with a random process of it's own.

Clearly, each location is correlated with other $n-1$ locations and $C(\theta) = C(-\theta)$ the variance-covariance matrix Σ is circulant. We use singular value decomposition (SVD) and obtain the correlated data $\underset{\sim}{X}$ on a circle as follows,

$$\underset{\sim}{X} = \Sigma^{1/2} * Z = Q\Lambda^{1/2}Q^T * Z,$$

where $\Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ and $Q = \{\psi_1, \psi_2, \cdots, \psi_n\}$ are eigen values and eigen vectors of the circulant matrix respectively and $Z \sim N(\underset{\sim}{0}, 1_n)$.

3.2.1 Covariance estimator comparison

In general the covarince estimator (3.1.1) on a circle is biased, with a bias of $var(\bar{X})$. In order to make things simple we set $C_1, a = 1$ and when $\alpha = 0.5$ $c_0 \geq \int_0^\pi (\theta)^{0.5} \sin\theta d\theta$, from Fresnel intergal it can be shown that $c_0 \geq 2.4353$. Now we compare the covariance estimator (empirical) to it's theoretical covariance given by 3.2.1 and 3.2.2. We computed the MOM estimator $\hat{C}(\theta)$ with 48 gridded observations on the circle from 500 simulations.

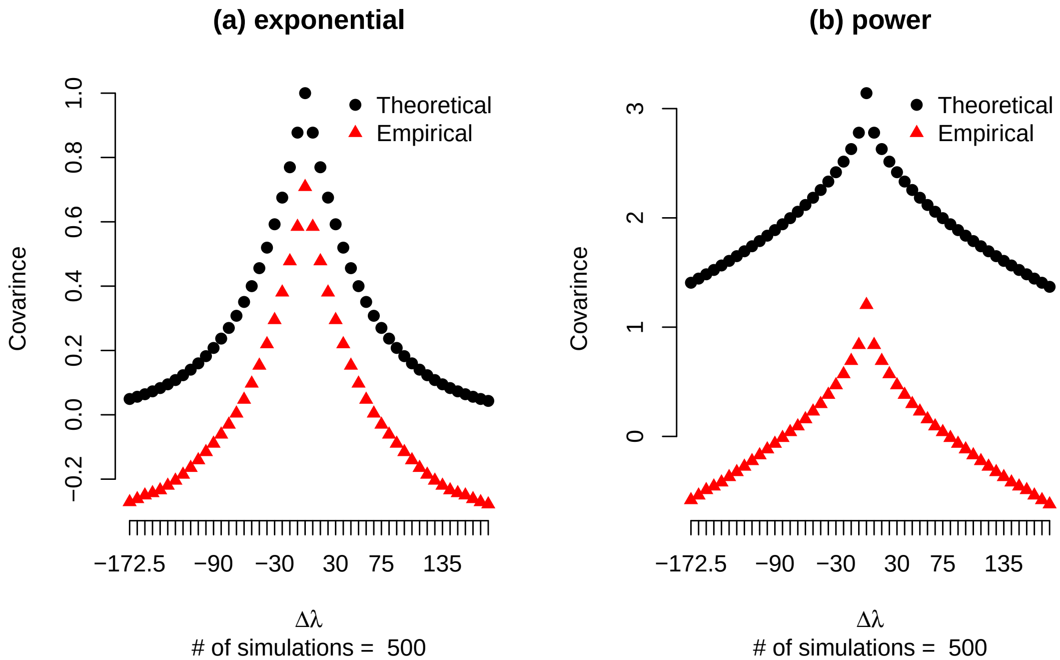

Figure 3.2: Theoretical and empirical covariance (with bias) comparison on a circle, it is easy to notice the bias in both covariance models

Remark 1: The shift between theoretical and empirical values is equal to a_0 and from $a_0 = \frac{1}{\pi}\int_0^{\pi} C(\theta)d\theta$ we can obtain

$$\begin{aligned} \text{exponential}: \; a_0 &= \frac{C_1}{a\pi}(1 - e^{-a\pi}) \\ \text{power}: \; a_0 &= c_0 - \left(\frac{\pi}{a}\right)^{\alpha} \frac{1}{\alpha + 1} \end{aligned}$$

Now consider the following covariance function, after subtracting a_0 from $C(\theta)$.

$$D(\theta) = C(\theta) - a_0.$$

If the new covariance function $D(\theta)$ was used to generate the data on a circle then the covariacne estimate converge to the theoretical value. In other words if the process is a zero mean process the covariance estimator given by 3.1.1 is unbiased $(i.e. Var(\bar{X}) = 0)$ hence we will get a perfect match between theoretical and empirical values.

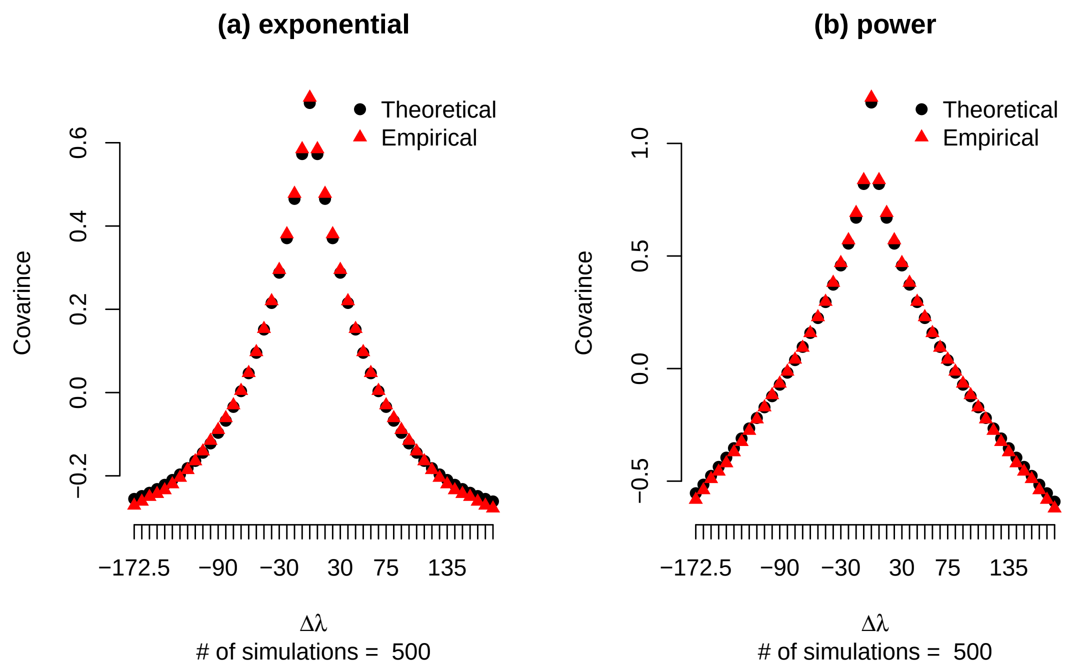

Figure 3.3: Theoretical and empirical covariance comparison on a circle using the covariance function $D(\theta)$.

Remark 2: The covariance estiamtor is biased and the biasness will approach to a_0. When covariance function is unknown the biaseness a_0 is also known and the biasness cannnot to estimated (cannot find the variance of $\bar{X}$) from one circle, however as discussed above if multiple *i.i.d.* copies on the same circle were available then one can estimate a_0 *i.e.* $\hat{a} = var(\bar{X})$ and substract $\hat{a_0}$ form the MOM estimator as given below,

$$\hat{C}(\Delta\lambda) = \left(\frac{1}{n} \sum_{i=1}^{n} (X(t_i + \Delta\lambda) - \bar{X})(X(t_i) - \bar{X}) \right) - \hat{a_0}$$

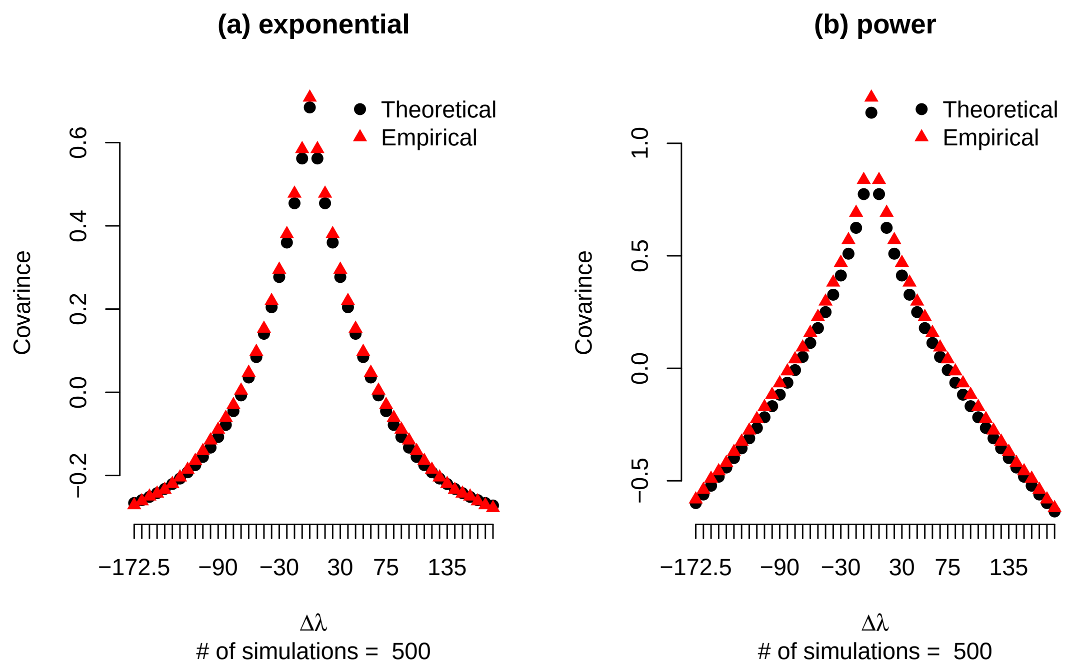

Figure 3.4: Theoretical and empirical covariance comparison on a circle, after removing $\hat{a}_0$ from the

3.2.2 Variogram estimator comparison

In general the variogram estimator in the case of a is unbiased but not consistent. When the random process on a circle is isotropy the semi variogram is given by

$$\gamma(\theta) = C(0) - C(\theta),$$

the theoretical variogram based on exponential and power covariance functions can be given in the following form,

$$\begin{aligned} \text{exponential}: \gamma(\theta) &= C(0) - C(\theta) = C_1(1 - e^{-a|\theta|}) \\ \text{power}: \gamma(\theta) &= C(0) - C(\theta) = (|\theta|/a)^{\alpha} \end{aligned}$$

sd

We computed the variogram estimator $\hat{\gamma}(\theta)$ with 48 gridded observations on the circle from 500 simulations since there is no bias varigram estimator is a better fit between theoretical and empirical values compared to covariance estimate.

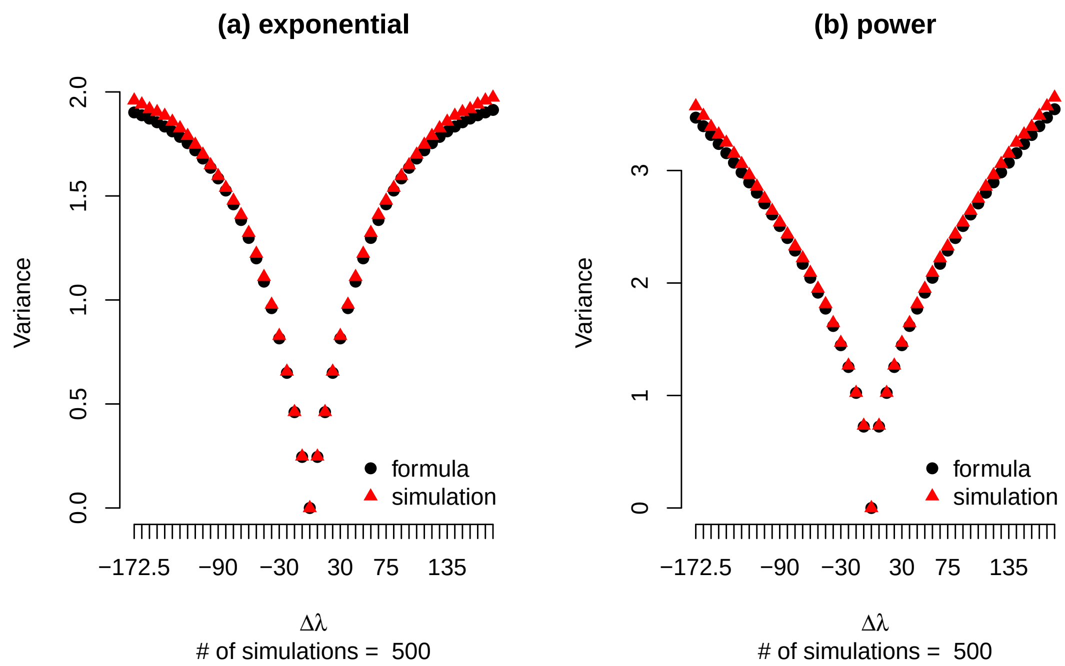

Figure 3.5: Theoretical and empirical comparison for variogram on a circle

Chapter 4

Parametric Models on a Sphere

4.1 Random process on a sphere

Suppose $X \in \{X(P) : P \in D\}$, defined in a common probability space $P \in S^2$ (unit sphere), where $P = (\lambda, \phi) \in S^2$ with longitude $\lambda \in [-\pi, \pi)$ and latitude $\phi \in [0, \pi]$. Suppose the process is isotropy and continuous in quadratic mean with respect to the location P then the process can be represented by spherical harmonics, $P_\nu^m(\cdot)$ normalized associated Legendre polynomials, with the sum converges in mean square (Jones (1963),Li and North (1997); Huang et al. (2012)).

$$X(P) = \sum_{\nu=0}^{\infty} \sum_{m=-\nu}^{\nu} Z_{\nu,m} e^{im\lambda} P_\nu^m(\cos\phi),$$

Since $\cos(\phi) \in [-1, 1]$ we have $\int_{-1}^{1} [P_\nu^m(cos(\phi))]^2 dcos(\phi) = 1$, and $Z_{\nu,m}$ are complex-valued coefficients satisfying.

$$Z_{\nu,m} = \int_{S^2} X(P) e^{-im\lambda} P_\nu^m(\cos\phi) dP.$$

Suppose the process $X(P)$ is isotropy with 0 mean (without loss of generality) which implies $E(Z_{\nu,m}) = 0$. Let $P = (\lambda_P, \phi_P)$ and $Q = (\lambda_Q, \phi_Q)$ be two arbitrary locations on the sphere, if the covariance function $R(P, Q)$ on S^2 solely depends on the spherical distance between (P, Q),

$$\begin{aligned} R(P,Q) &= \mathrm{E}(X(P)\overline{X(Q)}) \\ &= \sum_{\nu=0}^{\infty} \sum_{\mu=0}^{\infty} \sum_{m=-\nu}^{\nu} \sum_{n=-\mu}^{\mu} \mathrm{E}(Z_{\nu,m}\overline{Z}_{\mu,n}) e^{im\lambda_P} P_\nu^m(\cos\phi_P) e^{-in\lambda_Q} P_\mu^n(\cos\phi_Q). \end{aligned}$$

where $\bar{Z}$ denotes the complex conjugate of Z, θ_{PQ} is the spherical distance, $P_\nu(\cdot)$ is the Legendre polynomial. Note that the continuity of $X(P)$ on every point P implies that $R(P, Q)$ is continuous at all pairs of (P, Q) (Leadbetter, 1967, page 83).

Under the assumption of homogeneity (or isotropy) the random process $X(\cdot)$ on S^2 has a finite second moment and it is invariant under the rotations on the sphere with constant mean. Similarly, we can define an isotropic random process on a sphere as,

$$\begin{aligned} E(X(s) &= \mu \quad \text{for any } s \in S^2 \\ Cov(X(P), X(Q)) &= C(\theta_{PQ}) \end{aligned}$$

where θ_{PQ} is the spherical angle between two locations P, Q. For a unit sphere, the distance between the two locations can be defined as great circle distance (gcd_{PQ}) or chordal distance (ch_{PQ}) as follows,

$$\theta_{PQ} = \arccos\left(\sin(L_1)\sin(L_2) + \cos(L_1)\cos(L_2)\cos(l_1 - l_2)\right)$$

In the case of $\mathbb{R}^d$, non-negative definite is a necessary and sufficient condition for a valid covariance function defined on $\mathbb{R}^d$ (1.1.4). Similarly, a real continuous function $C(\cdot)$ defined on the sphere is a valid covariance function if and only if $C(\cdot)$ is non-negative definite,

$$\sum_{i,j=1}^{N} a_i a_j C(\theta_{PQ}) \geq 0, \tag{4.1.1}$$

for any integer N, any constants $a_1, a_2, \ldots, a_N$, and any locations $P, Q, \ldots \in S^2$.

Let $P_k^{\nu}(\cos\theta)$ be the ultraspherical polynomials defined by the following infinite summation,

$$\frac{1}{(1 - 2c\cos\theta + c^2)^{\nu}} = \sum_{k=0}^{\infty} c^k P_k^{\nu}(\cos\theta) \quad \nu > 0 \tag{4.1.2}$$

When $\nu = 0$, $P_k^0(\cos\theta) = \cos(k\theta)$

According to Schoenberg (1942), a real continuous function $C(\theta)$ is a valid covariance function on S^d, where $d = 1, 2, \ldots$, if and only if it can be written in the following form

$$C(\theta) = \sum_{k=0}^{\infty} c_k P_k^{(\nu)}(\cos\theta), \quad \nu = \frac{1}{2}(d-1) \tag{4.1.3}$$

where $\forall c_k \geq 0$ and $\sum c_k < \infty$.

Suppose $C(\cdot)$ is a covariance functions that is valid in S^d then it is valid on S^m where $d \leq m$. In general we have the following property,

$$\begin{gathered} S^1 \subset S^2 \subset \cdots S^d \subset \cdots S^{\infty} \\ C(S^1) \supset C(S^2) \supset \cdots C(S^d) \supset \cdots C(S^{\infty}) \end{gathered}$$

and covariance functions that are valid on S^m may not be valid on S^d where $m > d$. *should we add any examples?*

In chapter 3 we discussed about the spectral representation of the covariance function on a circle

$$C(\theta) = \sum_{k=0}^{\infty} c_k \cos(k\theta)$$

It is easy to observe that $\cos\theta \in S^1$ and clearly $\cos\theta \in S^2$ and from the properties of covariance discussed in chapter 1, $P_k \cos\theta \in S^2$ where P_k is a Legender polynomial. Therefore, when $d = 2$ the covarince on a sphere (S^2) can be expressed as follows,

$$C(\theta) = \sum_{k=0}^{\infty} c_k P_k(\cos\theta) \quad , c_k \geq 0 \tag{4.1.4}$$

Since the Legendre polynomials are orthogonal we have

$$\int_{-1}^{1} P_n(x)P_m(x)dx = \frac{2}{2n+1}\delta_{nm}$$

and on a sphere the coefficients c_k can be obtained by

$$c_\nu = \frac{2\nu+1}{2}\int_0^\pi C(\theta)P_\nu(\cos\theta)d\theta. \quad \nu = 0, 1, 2, \ldots \tag{4.1.5}$$

One can directly use the above integral to evaluate the validity of a covariance function on the sphere by checking if c_k is non-negative and $\sum c_k < \infty$.

All covariance models that are valid on $\mathbb{R}^d$ are not valid on the sphere (S^2), Huang et al. (2011) evaluated the validity of commonly used covariance on a sphere that are valid on $\mathbb{R}^d$, they showed that some models are not valid on the sphere and some models are valid only for certain parameter values.

Model	Covariance function	Validity S^2
Spherical	$\left(1 - \frac{3\theta}{2a} + \frac{1}{2}\frac{\theta^3}{a^3}\right) \mathbf{1}_{(\theta \leq a)}$	Yes
Stable	$\exp\left\{-\left(\frac{\theta}{a}\right)^\alpha\right\}$	Yes for $\alpha \in (0,1]$ No for $\alpha \in (1,2]$
Exponential	$\exp\{-\left(\frac{\theta}{a}\right)\}$	Yes
Gaussian	$\exp\left\{-\left(\frac{\theta}{a}\right)^2\right\}$	No
Power*	$c_0 - (\theta/a)^\alpha$	Yes for $\alpha \in (0,1]$ No for $\alpha \in (1,2]$
Radon transform of order 2	$e^{-\theta/a}(1 + \theta/a)$	No
Radon transform of order 4	$e^{-\theta/a}(1 + \theta/a + \theta^2/3a^2)$	No
Cauchy	$(1 + \theta^2/a^2)^{-1}$	No
Hole - effect	$\sin a\theta/\theta$	No

Table 4.1: Validity of covariance functions on the sphere, $a > 0, \theta \in [0, \pi]$. *When $\alpha \in (0, 1]$, power model is valid on the sphere for some $c_0 \geq \int_0^\pi (\theta/a)^\alpha \sin\theta d\theta$.

Furthermore, Gneiting (2013) argued that Matérn covariance function is valid on the sphere when the smoothness parameter $\nu \in (0, 1/2)$ and it is not valid if $\nu > 1/2$. Yadrenko (1983) showed that if $K(\cdot)$ is valid isotropic covariance function on $\mathbb{R}^3$ then

$$C(\theta) = K(2\sin(\theta/2))$$

is a valid isotropic covariance function on the unit sphere (where θ is *gcd*).

Similar to the case of circle, if a random process $X(\cdot)$ on a sphere is intrinsically stationary on S^2, then one has $E(X(P)) = \mu$ an unknown constant for all $P \in S^2$ and the variogram function between any two locations $P, Q \in S^2$ depends only on the spherical angle θ_{PQ}

$$Var(X(P) - X(Q)) = 2\gamma(\theta_{PQ}) \quad , \forall P, Q \in S^2$$

The variogram is conditionally negative definite

$$\sum_{i,j=1}^{N} a_i a_j 2\gamma(\theta_{PQ}) \leq 0,$$

for any integer N, any constants $a_1, a_2, \ldots, a_N$ with $\sum a_i = 0$, and any locations $P, Q, \ldots \in S^2$. Immediately from 4.1.4 for a continuous $2\gamma(\cdot)$ with $\gamma(0) = 0$ the variogram is negative definite if and only if

$$\gamma(\theta) = \sum_{k=0}^{\infty} c_k(1 - P_k(\cos\theta)) \tag{4.1.6}$$

where $P_k(\cdot)$ are Legendre polynomials with $\forall c_k \geq 0$ and $\sum c_k < \infty$.

In the introduction we pointed out in $\mathbb{R}^d$ one can always construct the variogram if the covariance function is given but not the converse. However, in S^2 Yaglom (1961) argued that for a valid $\gamma(\theta) \quad \theta \in [0, \pi]$ one can always construct covariance function $C(\theta) = c_0 - \gamma(\theta)$ for some $c_0 \geq \int_0^{\pi} \gamma(\theta)\sin(\theta)d\theta$.

4.2 Axially symmetry

All covariance models that are valid on R^3 are not valid on S^2 and it is necessary to use S^2 instead of R^3 when studying about random processes on Earth and isotropy is often assumed (Yadrenko (1983); Yaglom (1987)). However, many studies have pointed out this assumption is not reasonable (Stein (2007); Jun and Stein (2008); Bolin and Lindgren (2011)) for random processes on the sphere primarily on Earth. Stein (2007) argued that Total Ozone Mapping Spectrometer (TOMS) data varies strongly with latitudes and homogeneous models are not suitable. Moreover, aerosol depth (AOD) from Multi-angle Imaging Spectrometer (MISR), Sea Surface Temperature (SST) from RRMM Microwave Imager (TMI) are some other example for anisotropy global data on a sphere (on Earth). In order to study non homogeneous processes on the sphere Jones (1963) introduces the concept of axially symmetry, where the covarince between two spatial points depend on the longitudes only through their difference between two points.

A random process $X(P) : P \in S^2$ on the sphere and let $R(P, Q)$ be a valid covarince function on the sphere where $P = (L_P, l_P), Q(L_Q, l_Q)$ then $X(P)$ is axially symmetric if and only if

$$R(L_P, L_Q, l_P, l_Q) = R(L_P, L_Q, l_P - l_Q).$$

Currently, to our knowledge there are no methods to test axially symmetry in real data. However, this assumption is more plausible and reasonable when modeling spatial data. For example, temperature, moisture, etc. most likely symmetric on longitudes rather than latitudes. Stein (2007) propose a method to model axially symmetric process on a sphere (the fitted model is not the best, but this was a good start). When modeling spatial data stationary models are less useful; but using the concept of axially symmetry Jun and Stein (2008) proposed a flexible class of parametric covariance models to capture the non-stationarity of global data. Hitczenko and Stein (2012) discussed about the properties of an existing class of models for axially symmetric Gaussian processes on the sphere. They applied first-order differential operators to an isotropic process. Huang et al. (2012) developed a new representation of axially symmetric process on the sphere and further introduced some parametric covariance models that are valid on S^2.

if the process is axially symmetric $E(Z_{\nu,m}\overline{Z}_{\mu,n})$ can be expressed as,

$$\mathrm{E}(Z_{\nu,m}\overline{Z}_{\mu,n}) = \delta_{n,m} f_{\nu,\mu,m}.$$

Hence, for an axially symmetric process the covariance function (4.1.1) will be the following form (Huang et al. (2012))

$$\begin{aligned} R(P,Q) &= R(\phi_P, \phi_Q, \lambda_P - \lambda_Q) \\ &= \sum_{m=-\infty}^{\infty} \sum_{\nu=|m|}^{\infty} \sum_{\mu=|m|}^{\infty} f_{\nu,\mu,m} e^{im(\lambda_P - \lambda_Q)} P_\nu^m(\cos\phi_P) P_\mu^m(\cos\phi_Q). \end{aligned} \tag{4.2.1}$$

In order to have a valid covariance function, $f_{\nu,\mu,m} = \overline{f}_{\mu,\nu,m}$ and for each fixed integer m, the matrix $F_m(N) = \{f_{\nu,\mu,m}\}_{\nu,\mu=|m|,|m|+1,\ldots,N}$ must be positive definite for all $N \geq |m|$.

$$R(P,Q) = R(\phi_P, \phi_Q, \Delta\lambda) = \sum_{m=-\infty}^{\infty} e^{im\Delta\lambda} C_m(\phi_P, \phi_Q) \quad m = 0, \pm 1, \pm 2, ... \tag{4.2.2}$$

where $\Delta\lambda \in [-\pi, \pi]$ and $\phi_P, \phi_Q \in [0, \pi]$

4.2.1 Properties of $C_m(\phi_P, \phi_Q)$

The covariance function $R(P,Q)$ based on the concept of axially symmetry is clearly defined by both latitudes and longitudes (difference). The following conditions for $C_m(\phi_P, \phi_Q)$ are very important to have a valid covariance function defined by 4.2.2.

- Hermitian and positive definite.
- $\sum_{m=-\infty}^{\infty} |C_m(\phi_P, \phi_Q)| < \infty$ for $m = 0, \pm 1, \pm 2, ...$
- Is a continuous function.

One can use inverse Fourier transformation to derive C_m based on an axially symmetric covariance function $R(P,Q)$ defined on a sphere, as we have

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

Let's consider a real-valued process with a complex valued $C_m(\phi_P, \phi_Q)$ as given below,

$$\begin{aligned} C_m(\phi_P, \phi_Q) &= c_m f(\phi_P, \phi_Q) e^{i\omega_m(\phi_P - \phi_Q)}, \quad c_m \geq 0, \omega_m \in R \\ &= c_m C_m^R(\phi_P, \phi_Q) + i c_m C_m^I(\phi_P, \phi_Q). \end{aligned}$$

Huang et al. (2012) states that if a process is real-valued then the corresponding covariance function $R(P,Q)$ is also real-valued and $C_m(\phi_P, \phi_Q) = \overline{C_{-m}(\phi_P, \phi_Q)}$, The covarince function $R(P,Q)$ on the sphere given by 4.2.2 can be simplified to the following form,

$$\begin{aligned} R(P,Q) &= C_0(\phi_P,\phi_Q) + \sum_{m=1}^{\infty} e^{-im\Delta\lambda} C_{-m}(\phi_P,\phi_Q) + \sum_{m=1}^{\infty} e^{im\Delta\lambda} C_m(\phi_P,\phi_Q) \\ &= c_o C_0^R(\phi_P,\phi_Q) + 2\sum_{m=1}^{\infty} c_m[\cos(m\Delta\lambda) C_m^R(\phi_P,\phi_Q) - \sin(m\Delta\lambda) C_m^I(\phi_P,\phi_Q)]. \end{aligned}$$

There are several covariance models, $R(P,Q)$, valid on a sphere suggested by Huang et al. (2012) carefully choosing values for c_m.

Model	c_m	parameters
model 1	: $c_m = Cp^m$	$m = 0, \pm 1, \pm 2, \ldots \quad p \in (0,1)$
model 2	: $c_m = \frac{Cp^m}{m}$ and $c_0 = 0$	$m = \pm 1, \pm 2, \ldots \quad p \in (0,1)$
model 3	: $c_m = \frac{C}{m^4}$ and $c_0 = 0$	$m = \pm 1, \pm 2, \ldots$

Table 4.2: some proposed c_m models

Suppose $K(\cdot)$ is a valid covariance function defined on a sphere where,

$$K(L_1, L_2, l_1 - l_2 = K(L_1, L_2, l_2 - l_1) \tag{4.2.3}$$

is special a case for axially symmetric process and the underline process is said to be longitudinally reversible, the idea was first introduced by Stein (2007).

For example the covariance model proposed by Huang et al. (2012) clearly yields a longitudinally reversible process as $R(\phi_P, \phi_Q, \Delta\lambda) = R(\phi_P, \phi_Q, -\Delta\lambda)$ and the reversibility holds when $C_{-m}(\phi_P, \phi_Q) = C_m(\phi_P, \phi_Q)$. Now the covariance function reduces to the following,

$$R(P,Q) = \sum_{m=0}^{\infty} C_m(\phi_P,\phi_Q)\cos(m\Delta\lambda)$$

If a random process on the sphere is real valued and longitudinally reversible so is the covariance function, $R(P,Q)$, is real valued then $C_m(\phi_P, \phi_Q)$ is real since $C_{-m}(\phi_P, \phi_Q) = C_m(\phi_P, \phi_Q)$ and $C_m(\phi_P, \phi_Q) = \overline{C_{-m}(\phi_P, \phi_Q)}$.

4.3 Generalization of parametric models

How to discuss/link about other parametric models suggested by Jeong and Jun (2015), Jun and Stein (2008), etc... and why are we using Huang et al. (2012) and what are the advantages?

The covariance function on sphere, $R(P,Q)$, given in equation 4.2.2, is clearly a function of both longitude and latitude.

$$R(P,Q) = f(\Delta\lambda, \phi_P, \phi_Q)$$

In order to make things easier one could assume that $C_m(\phi_P, \phi_Q) = \tilde{C}_m(\phi_P - \phi_Q)$ only depends on the difference of ϕ_P and ϕ_Q, Huang et al. (2011) proposed a simple separable covariance function when both covariance components are exponential

$$R(P, Q) = c_0 e^{-a|\Delta\lambda|} e^{-b|\phi_P - \phi_Q|},$$

Where a and b are defined as decay parameters in longitude and latitude respectively. The separable models are too simple and they are not capable to capture the covariance structure of the entire sphere. Therefore, Huang et al. (2012) proposed some non-separable covariance models by carefully choosing functions for $C_m(\phi_P, \phi_Q)$ that are valid on the sphere,

$$R(P, Q) = Ce^{-a|\phi_P - \phi_Q|} \frac{1 - p^2}{1 - 2p\cos\Theta + p^2} \quad (4.3.1)$$

$$R(P, Q) = Ce^{-a|\phi_P - \phi_Q|} \log \frac{1}{(1 - 2p\cos\Theta * + p^2)} \quad (4.3.2)$$

$$R(P, Q) = 2Ce^{-a|\phi_P - \phi_Q|} \left(\frac{\pi^4}{90} - \frac{\pi^2\Theta^2}{12} + \frac{\pi\Theta^3}{12} - \frac{\Theta^4}{48} \right), \quad (4.3.3)$$

where $\Theta = \Delta\lambda + u(\phi_P - \phi_Q) - 2k\pi$, and k is chosen such that $\Theta \in [0, 2\pi]$.

There is one big disadvantage of the covariance models proposed by Huang et al. (2012), the biggest disadvantage for all of them are that it is assumed not only stationarity on longitudes, but stationarity on latitudes as well.

We have noticed that when $\phi_P = \phi_Q$, the model 4.3.1 reduces to

$$R(P, P) = C \frac{1 - p^2}{1 - 2p\cos(\Delta\lambda) + p^2}$$

and if we set $\Delta\lambda = 0$, the variance of latitude ϕ_P over all latitudes can be given by,

$$Var(P) = C \frac{1 + p}{1 - p}$$

is not a function of the latitude (a function of the parameter p) and it implies that variance is constant over all latitudes. This is not supposed to be the case, since both MSU data and TOMS data in figures 2.4 and 2.7 shows that variance is highly depending on the latitude. In order to overcome this issue we propose a solution to the above covariance models to capture the non stationarity over the latitudes.

Proposition 4.3.1 *A more general non stationary covariance function is given as following. If $C(\cdot) = C(x - y)$ is the stationary covariance function and $f(\omega) \geq 0$ is the corresponding spectral density, then*

$$\tilde{C}(x, y) = C_2 - C(x) - C(y) + C(x - y),$$

with

$$C_2 \geq \int_{-\infty}^{\infty} dF(\omega) = \int_{-\infty}^{\infty} f(\omega)d\omega > 0$$

is the non stationary covariance function. Note that the covariance function $C(\cdot)$ implies that, by Bochner's theorem, there exists a bounded measure F such that

$$C(x) = \int_{-\infty}^{\infty} e^{-ix\omega} dF(\omega).$$

When $F(\cdot)$ is absolutely continuous, there exists a spectral density $f(\cdot) \geq 0$ such that

$$C(x) = \int_{-\infty}^{\infty} e^{-ix\omega} f(\omega)d\omega.$$

Now we choose a sequence of complex numbers $a_i, i = 1, 2, \cdots, n$, and any sequence of real numbers $t_i, i = 1, 2, \cdots, n$, taking $C_2 = \int_{-\infty}^{\infty} f(\omega)d\omega$,

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\overline{a}_j\tilde{C}(t_i, t_j) &= \sum_{i}\sum_{j} a_i\overline{a}_j(C_2 - C(t_i) - C(-t_j) + C(t_i - t_j)) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\overline{a}_j \int_{-\infty}^{\infty}(1 - e^{-it_i\omega} - e^{it_j\omega} + e^{-i(t_i-t_j)\omega})f(\omega)d\omega \\
&= \int_{-\infty}^{\infty} f(\omega)d\omega \left|\sum_{i=1}^{n} a_i(e^{-it_i\omega} - 1)\right|^2 \geq 0.
\end{aligned}$$

In the case of circle clearly $C_1 e^{|\theta|}$ is a stationary covariance function and we can apply 4.3.1 to get a non stationary covariance function. Lets consider the following stationary covariance functions over the latitudes,

$$C(\phi) = Ce^{-a|\phi_P|}$$
$$C(\phi) = C\frac{1}{\sqrt{a^2 + \phi^2}}$$

Now, we apply proposition 4.3.1 to get non-stationary covariance functions, which depends on the latitudes, even when $\phi_P = \phi_Q$. Consider the below two functions for $C_m(\phi_P, \phi_Q)$.

$$\tilde{C}(\phi_P, \phi_Q) = C_1(C_2 - e^{-a|\phi_P|} - e^{-a|\phi_Q|} + e^{-a|\phi_P - \phi_Q|}) \tag{4.3.4}$$

$$\tilde{C}(\phi_P, \phi_Q) = C_1\left(C_2 - \frac{1}{\sqrt{a^2 + \phi_P^2}} - \frac{1}{\sqrt{a^2 + \phi_Q^2}} + \frac{1}{\sqrt{a^2 + (\phi_P - \phi_Q)^2}}\right) \tag{4.3.5}$$

Here $C_1, a > 0$, and $C_2 \geq 1$ to ensure the positive definiteness of the above function. When $\phi_P = \phi_Q$, both functions are actually a function of ϕ_P.

$$
\begin{aligned}
\tilde{C}(\phi_P, \phi_P) &= C_1(C_2 - 2e^{-a|\phi_P|} + 1), \\
\tilde{C}(\phi_P, \phi_P) &= C_1\left(C_2 - \frac{2}{\sqrt{a^2 + \phi_P^2}} + \frac{1}{a}\right).
\end{aligned}
$$

So we propose six five-parameter models which are combinations of both $\tilde{C}(\phi_P, \phi_Q)$, defined by a exponential family 4.3.4 and a power family 4.3.5, and models (4.3.1, 4.3.2, 4.3.3) proposed by Huang et al. (2012) for the covariance on a sphere defined as follows,

$$
R(P, Q) = \tilde{C}(\phi_P, \phi_Q)C(\theta(P, Q, u)),
$$

where $\theta(P, Q, u) = \Delta\lambda + u(\phi_P - \phi_Q) \in [0, 2\pi]$, $C_1 > 0, C_2 > 0, a > 0, u \in \mathbb{R}, p \in (0, 1)$.

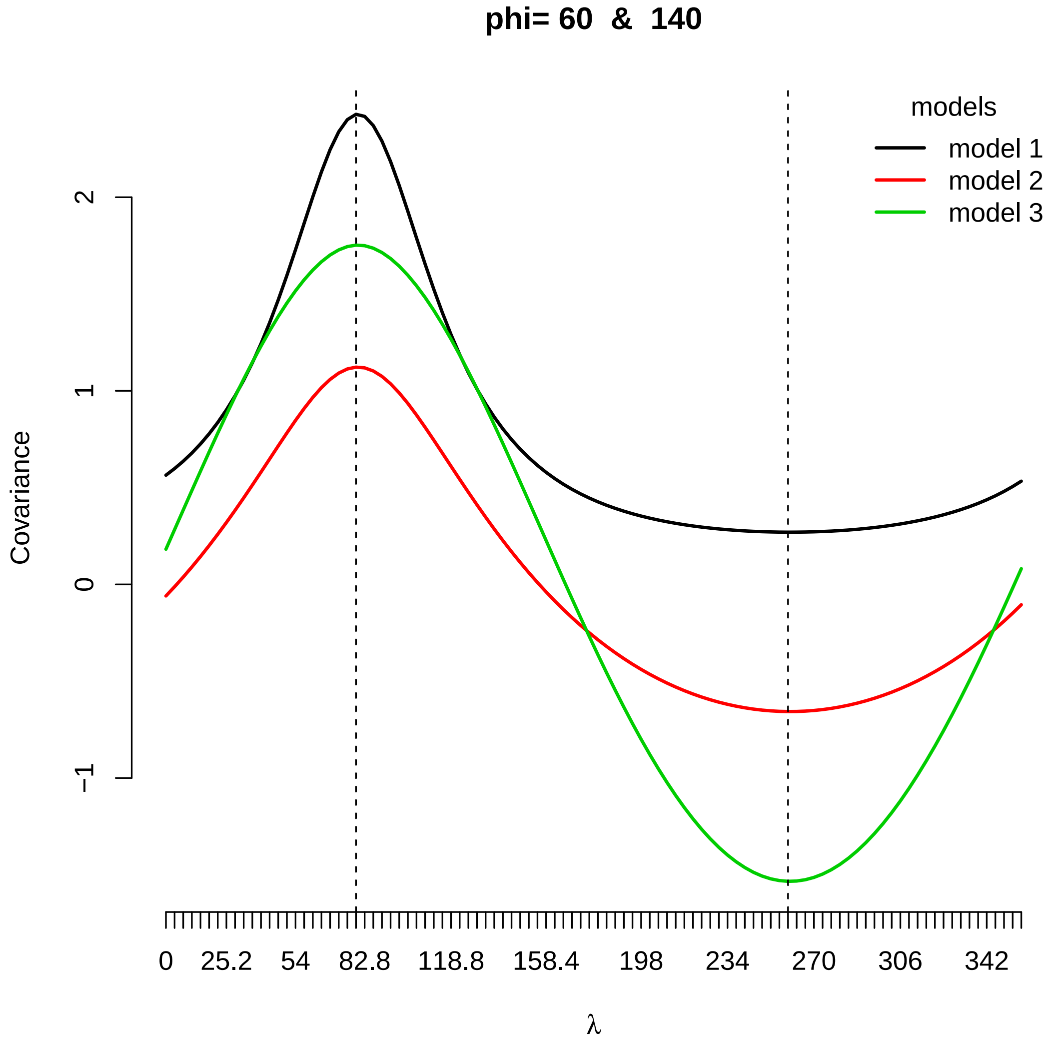

Figure 4.1: The covariance between 30^0S and 50^0N (latitude 60^0 and 140^0) of three covariance models with exponential family $i.e.\tilde{C}(\phi_P, \phi_Q)$ given by 4.3.4 over 100 longitudes for simplicity we set all parameters to be one.

Remark 1 The parameters C_1, C_2, a, p are scaling parameters of the covariance functions and u is a location parameter. All covariance models have a similar pattern and share one property, when there is no location shift ($u = 1$) the maximum of $R(P, Q)$ occurs at $\lambda_{max} = |\phi_P - \phi_Q|$ and the minimum of $R(P, Q)$ occurs at $\lambda_{min} = \pi + \lambda_{max}$.

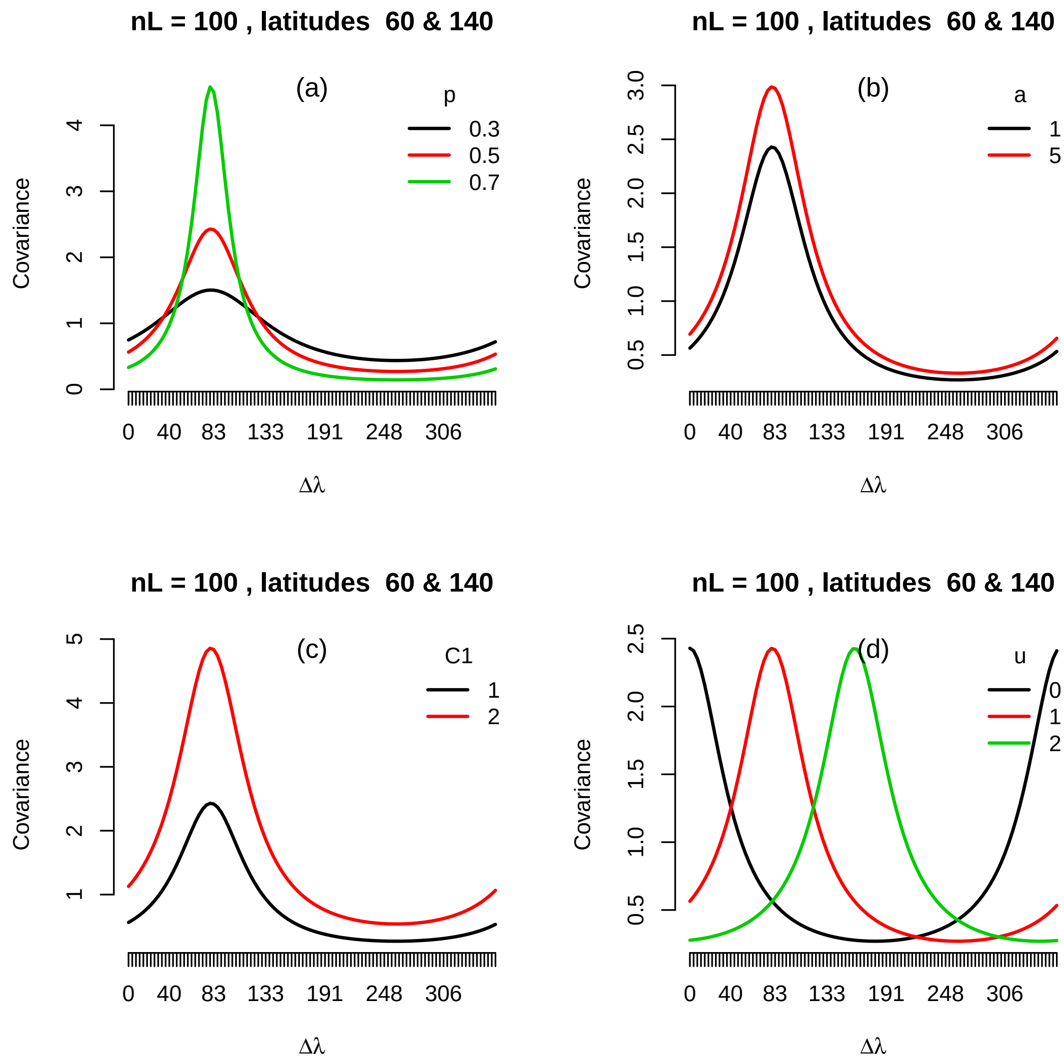

Figure 4.2: Covariance distribution for different parameter using model1: (a)-parameter p, (b)-parameter a, (c)-parameter $C1$ (similar pattern for parameter $C2$), (d)-parameter u

Remark 2 The scaling parameter p is more sensitive at supremum and infimum of the covariance models and parameters C_1, C_2, a are regular scaling parameters. The parameter u is a location parameter which shits the covariance from left to right $(\Delta\lambda)$ when $u > 0$, but $u = 0$ will provide a longitudinally reversible covariance model which is similar to Matérn covariance model when the smoothing parameter (ν) is 1/2.

4.4 Covariance and variogram estimators on a sphere

Cross Covariance

The cross covariance captures the covarince between two locations and any finite pairs of locations separated by a fixed distance (longitudinal difference $\Delta\lambda$). In other words cross covariance can be used to to capture the covariance between points at two latitudes separated by $\Delta\lambda \in (0, 2\pi)$. When a axially symmetric random process on a sphere is second-order stationary, cross covariance is a function of longitudinal difference ($\Delta\lambda$). According to Wackernagel (2013) the cross covariance function is not an even function and it is easy to observe that the proposed $R(P, Q)$ functions are valid on a sphere and they are cross covarince functions ($R(P, Q, \Delta\lambda) \neq R(P, Q, -\Delta\lambda)$). The cross covariance estimate for axially symmetric processes on the sphere. For any two latitudes ϕ_P and ϕ_Q with $\{\lambda_i, i = 1, 2, \ldots, n\}$ representing the gridded longitudes on each circle, then $\hat{R}(\phi_P, \phi_Q, \Delta\lambda)$ is given by

$$\hat{R}(\phi_P, \phi_Q, \Delta\lambda) = \frac{1}{n}\sum_{i=1}^{n}(X(\phi_P, \lambda_i + \Delta\lambda) - \bar{X}_P)(X(\phi_Q, \lambda_i) - \bar{X}_Q), \tag{4.4.1}$$

where $\Delta\lambda = 0, 2\pi/n, 4\pi/n, \cdots, 2(N-1)\pi/n$ and $\bar{X}_P = \frac{1}{n}\sum_{i=1}^{n} X(\phi_P, \lambda_i)$ and similar for $\bar{X}_Q$. Now we calculate the unbiasedness of the cross covarince estimator.

$$\begin{aligned}
E(\hat{R}(\phi_P, \phi_Q, \Delta\lambda)) &= \frac{1}{n}\sum_{i=1}^{n} E((X(\phi_P, \lambda_i + \Delta\lambda) - \bar{X}_P)(X(\phi_Q, \lambda_i) - \bar{X}_Q)) \\
&= \frac{1}{n}\sum_{i=1}^{n} cov(X(\phi_P, \lambda_i + \Delta\lambda), X(\phi_Q, \lambda_i)) \\
&\quad -\frac{1}{n}\sum_{i=1}^{n} E((X(\phi_P, \lambda_i + \Delta\lambda) - \mu_P)(\bar{X}_Q - \mu_Q)) \\
&\quad -\frac{1}{n}\sum_{i=1}^{n} E((X(\phi_Q, \lambda_i) - \mu_Q)(\bar{X}_P - \mu_P)) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n} E((\bar{X}_P - \mu_P)(\bar{X}_Q - \mu_Q)) \\
&= R(\phi_P, \phi_Q, \Delta\lambda) - E((\bar{X}_Q - \mu_Q)(\bar{X}_P - \mu_P)) - E((\bar{X}_P - \mu_P)(\bar{X}_Q - \mu_Q)) \\
&\quad +E((\bar{X}_P - \mu_P)(\bar{X}_Q - \mu_Q)) \\
&= R(\phi_P, \phi_Q, \Delta\lambda) - cov(\bar{X}_P, \bar{X}_Q).
\end{aligned}$$

Note that,

$$
\begin{aligned}
cov(\bar{X}_P, \bar{X}_Q) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} cov(X(\phi_P,\lambda_i), X(\phi_Q,\lambda_j)) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} R(\phi_P,\phi_Q,(i-j)*2\pi/n) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} (C_0(\phi_P,\phi_Q) \\
&\quad 2\sum_{m=1}^{\infty}(C_{m,R}(\phi_P,\phi_Q)\cos(m*(i-j)*2\pi/n) \\
&\quad -C_{m,I}(\phi_P,\phi_Q)\sin(m*(i-j)*2\pi/n)) \\
&= C_0(\phi_P,\phi_Q) + 2\sum_{m=1}^{\infty} C_{m,R}(\phi_P,\phi_Q)\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\cos(m(i-j)*2\pi/n)\right) \\
&\quad -2\sum_{m=1}^{\infty} C_{m,I}(\phi_P,\phi_Q)\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sin(m(i-j)*2\pi/n)\right) \\
&= C_0(\phi_P,\phi_Q),
\end{aligned}
$$

since

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\cos(m*(i-j)*2\pi/n) \\
= &\sum_{i=1}^{n}\sum_{j=1}^{n}(\cos(m*i*2\pi/n)\cos(m*j*2\pi/n) - \sin(m*i*2\pi/n)\sin(m*j*2\pi/n)) \\
= &\left(\sum_{i=1}^{n}\cos(m*i*2\pi/n)\right)^2 - \left(\sum_{i=1}^{n}\sin(m*i*2\pi/n)\right)^2 = 0
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\sin(m*(i-j)*2\pi/n) \\
= &\sum_{i=1}^{n}\sum_{j=1}^{n}(\sin(m*i*2\pi/n)\cos(m*j*2\pi/n) - \cos(m*i*2\pi/n)\sin(m*j*2\pi/n)) \\
= &\left(\sum_{i=1}^{n}\cos(m*i*2\pi/n)\right) * \left(\sum_{i=1}^{n}\sin(m*i*2\pi/n)\right) \\
&-\left(\sum_{i=1}^{n}\cos(m*i*2\pi/n)\right) * \left(\sum_{i=1}^{n}\sin(m*i*2\pi/n)\right) = 0
\end{aligned}
$$

since for any integer m, we have

$$\sum_{k=1}^{n} \cos(mk * 2\pi/n) = \begin{cases} 0, & \text{for any integer } m \neq 0, \\ n, & \text{for } m = 0 \end{cases} \quad \text{and} \sum_{k=1}^{n} \sin(mk * 2\pi/n) = 0.$$

Hence,

$$cov(\bar{X}_P, \bar{X}_Q) = C_0(\phi_P, \phi_Q).$$

Therefore,

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

The cross covariance estimator is biased and when $\phi_P = \phi_Q$ this reduces to the same results we obtained for a random process on the circle.

Cross variogram

In general for a stationary process when the covariance is known one can get the variogram ($2\gamma(\theta) = C(0) - C(\theta)$), since the cross covariance is not an even function the variogram is defined by taking the average of $R(P, Q, \Delta\lambda)$ and $R(P, Q, -\Delta\lambda)$ and we can derive the cross variogram as follows,

$$\begin{aligned}
\gamma(\phi_p, \phi_Q, \Delta\lambda) &= \frac{1}{2} E\left((X(\phi_P, \lambda + \Delta\lambda) - X(\phi_P, \lambda))(X(\phi_Q, \lambda + \Delta\lambda) - X(\phi_Q, \lambda))\right) \\
&= \frac{1}{2} E\left(((X(\phi_P, \lambda + \Delta\lambda) - \mu_P) - (X(\phi_P, \lambda) - \mu_P))\right. \\
&\quad \left.((X(\phi_Q, \lambda + \Delta\lambda) - \mu_Q) - (X(\phi_Q, \lambda) - \mu_Q)\right) \\
&= \frac{1}{2}\left(cov(X(\phi_P, \lambda + \Delta\lambda), X(\phi_Q, \lambda + \Delta\lambda)) - cov(X(\phi_P, \lambda + \Delta\lambda), X(\phi_Q, \lambda))\right. \\
&\quad \left. - cov(X(\phi_P, \lambda), X(\phi_Q, \lambda + \Delta\lambda)) + cov(X(\phi_P, \lambda), X(\phi_Q, \lambda))\right) \\
&= \frac{1}{2}\left(R(\phi_P, \phi_Q, 0) - R(\phi_P, \phi_Q, \Delta\lambda) - R(\phi_P, \phi_Q, -\Delta\lambda) + R(\phi_P, \phi_Q, 0)\right) \\
&= R(\phi_P, \phi_Q, 0) - \frac{1}{2}(R(\phi_P, \phi_Q, \Delta\lambda) + R(\phi_P, \phi_Q, -\Delta\lambda)).
\end{aligned}$$

$$\gamma(\phi_p, \phi_Q, \Delta\lambda) = R(\phi_p, \phi_Q, 0) - \frac{1}{2}(R(\phi_P, \phi_Q, \Delta\lambda) + R(\phi_P, \phi_Q, -\Delta\lambda)). \tag{4.4.2}$$

The MOM estimator for cross-variogram on axially symmetric processes on the sphere is given by

$$\hat{\gamma}(\phi_p, \phi_Q, \Delta\lambda) = \frac{1}{2n} \sum_{i=1}^{n} \left(X(\phi_P, \lambda_i + \Delta\lambda) - X(\phi_P, \lambda_i))(X(\phi_Q, \lambda_i + \Delta\lambda) - X(\phi_Q, \lambda_i))\right), \tag{4.4.3}$$

and we have

$$
\begin{aligned}
E(\hat{\gamma}_{PQ}(\Delta\lambda)) &= \frac{1}{2n}\sum_{i=1}^{n} E\left(X(\phi_P, \lambda_i+\Delta\lambda) - X(\phi_P, \lambda_i))(X(\phi_Q, \lambda_i+\Delta\lambda) - X(\phi_Q, \lambda_i)\right)) \\
&= \frac{1}{2n}\sum_{i=1}^{n} (2\gamma(\phi_p, \phi_Q, \Delta\lambda)) = \gamma(\phi_p, \phi_Q, \Delta\lambda),
\end{aligned}
$$

which is unbiased.

Remark 3 Lets re arrange $R(P, Q)$,

$$
R(\phi_P, \phi_Q, \Delta\lambda) = \frac{1}{2}(R(\phi_P, \phi_Q, \Delta\lambda) + R(\phi_P, \phi_Q, -\Delta\lambda)) + \frac{1}{2}(R(\phi_P, \phi_Q, \Delta\lambda) - R(\phi_P, \phi_Q, -\Delta\lambda))
$$

the cross-covariance function $R(\phi_P, \phi_Q, \Delta\lambda)$ is decomposed into two components: the even component (the first average) and the odd component (the second average). The cross-variogram is only related to the even component of the cross-covariance function, which is different from the case on the circle (the covariance is an even function). Wackernagel (2013) argues that cross variogram is not sufficient when there is a delayed affect. However, in the data generation process they is no delayed affect (between latitudes).

Chapter 5

Global Data Generation and Estimation on the Sphere

The global data generation process on a sphere discussed on this dissertation is primarily based on the axially symmetric covariance structure introduced by Jones (1963) and as continuation of axially symmetric process on a sphere developed by Huang et al. (2012). Let $X(P)$ be a complex-valued random process defined on a unit sphere S^2, where $P = (\lambda, \phi) \in S^2$ with longitude $\lambda \in [-\pi, \pi)$ and latitude $\phi \in [0, \pi]$. In chapter 4 we discussed how to formulate a valid covariance function for continuous axially symmetric processes on a sphere and was given by 4.2.2. Now, in the light of Huang et al. (2012)[remark 2.5] a continuous axially symmetric process, $X(P)$ on a unit sphere , is given by

$$X(P) = X(\phi, \lambda) = \sum_{m=-\infty}^{\infty} W_{m,\nu}(\phi) e^{im\lambda} \psi_{m,\nu}(\phi), \tag{5.0.1}$$

where λ is the longitude, ϕ is the latitude and $\psi_{m,\nu}(\cdot)$ is a orthonormal basis for $C_m(\phi_P, \phi_Q)$ and using inverse Fourier transformation we get

$$W_m(\phi) = \frac{1}{2\pi} \int_{S^2} X(P) e^{-im\lambda} \overline{\psi_{m,\nu}(\phi)} dP,$$

with $\mathrm{E}(W_{m,\nu}\overline{W_{n,\nu}}) = \delta_{m,n}\delta_{m,\mu}\eta_{m,\nu}$.

Remark 1 If the random process on a sphere is real-Gaussian, then the weights given by $W_{m,\nu}$ will be independent normal random variables. Furthermore, if ν is fixed the process defined by $X(P)$ will be equivalent to a homogeneous random process on a circle with angular distance $\Delta\lambda$. In other words a random process on a sphere at a given latitude (for fix ϕ) can be studies as a random process on the circle. In Chapter 3 a random process on a circle was given by a infinite Fourier summation (Roy (1972),Dufour and Roy (1976)) and in the case of a process on a circle, $X(P)$ (5.0.1) can be given by

$$X(\phi, \lambda) = \sum_{m=-\infty}^{\infty} W_m(\phi) e^{im\lambda}, \tag{5.0.2}$$

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

with $\mathrm{E}(W_m(\phi_P)\overline{W_n(\phi_Q)}) = \delta_{m,n}C_m(\phi_P,\phi_Q)$.

5.1 Method development

We can construct normal independent (complex) random variate $W_m(\phi)$ associated with the variance-covariance matrix $C_m(\phi_P,\phi_Q)$ to construct an axially symmetric process for a given latitude ϕ. Then finite summation can be used to approximate above (5.0.2) infinite summation as given below,

$$X(P) = X(\phi,\lambda) = \sum_{m=-N}^{N} W_m(\phi)e^{im\lambda} \tag{5.1.1}$$

where this would provide the gridded data. Since W_m's are independent for $m = 1, 2, \cdots$, we have

$$\begin{aligned} Cov(X(P), X(Q)) &= Cov\left(\sum_{m=-N}^{N} W_m(\phi_P)e^{im\lambda_P}, \sum_{j=-N}^{N} W_j(\phi_Q)e^{ij\lambda_Q}\right) \\ &= \sum_{m,j} e^{im\lambda_P}e^{-ij\lambda_Q}Cov(W_m(\phi_P), W_j(\phi_Q)) \\ &= \sum_{m} e^{im(\lambda_P-\lambda_Q)}C_m(\phi_P,\phi_Q) \end{aligned}$$

The above generated data will be complex random variates. Therefore to have the real-valued data observations or to obtain a real process, we need to have

$$C_{-m}(\phi_P,\phi_Q) = \overline{C_m(\phi_P,\phi_Q)}, \quad \text{for } m = 1, 2, \cdots, N \tag{5.1.2}$$

Lets write $W_m(\phi) = W_m^r(\phi) + iW_m^i(\phi)$ in terms of a real component and an imaginary component. We also write $C_m(\phi_P,\phi_Q) = C_m^r(\phi_P,\phi_Q) + iC_m^i(\phi_P,\phi_Q)$ and with the relationship 5.1.2 above, we have

$$C_{-m}^r(\phi_P,\phi_Q) = C_{-m}^r(\phi_P,\phi_Q), \quad C_{-m}^i(\phi_P,\phi_Q) = -C_{-m}^i(\phi_P,\phi_Q).$$

Now,

$$\begin{aligned} Cov(W_m(\phi_P), W_m(\phi_Q)) &= Cov(W_m^r(\phi_P) + iW_m^i(\phi_P), W_m^r(\phi_Q) + iW_m^i(\phi_Q)) \\ &= \left[Cov(W_m^r(\phi_P), W_m^r(\phi_Q)) + Cov(W_m^i(\phi_P), W_m^i(\phi_Q))\right] \\ &\quad +i\left[-Cov(W_m^r(\phi_P), W_m^i(\phi_Q)) + Cov(W_m^i(\phi_P), W_m^r(\phi_Q))\right] \\ &= C_m^r(\phi_P,\phi_Q) + iC_m^i(\phi_P,\phi_Q). \end{aligned}$$

If we let $W_{-m}(\phi) = \overline{W_m(\phi)}$, then the covariance function would satisfy the above relationship 5.1.2. In addition, we will set the following,

$$Cov(W_m^r(\phi_P), W_m^r(\phi_Q)) = Cov(W_m^i(\phi_P), W_m^i(\phi_Q)) = \frac{1}{2}C_m^r(\phi_P, \phi_Q), \qquad (5.1.3)$$

$$Cov(W_m^i(\phi_P), W_m^r(\phi_Q)) = -Cov(W_m^r(\phi_P), W_m^i(\phi_Q)) = \frac{1}{2}C_m^i(\phi_P, \phi_Q). \qquad (5.1.4)$$

Therefore, if we denote $\underset{\sim}{W}_m(\phi) = (W_m^r(\phi), W_m^i(\phi))^T$, then the variance-covariance matrix for $\underset{\sim}{W}_m(\phi)$ is given by

$$\frac{1}{2}\begin{pmatrix} C_m^r(\phi_P, \phi_Q) & -C_m^i(\phi_P, \phi_Q) \\ C_m^i(\phi_P, \phi_Q) & C_m^r(\phi_P, \phi_Q) \end{pmatrix}.$$

However, we cannot have a vector of random variables $\underset{\sim}{W}_m(\phi)$ with a non-symmetric variance-covariance matrix unless $C_m^i(\phi_P, \phi_Q) = 0$. In the next section we will demonstrate how to generate $\underset{\sim}{W}_m(\phi)$ with a symmetric variance-covariance

The process given by (5.0.2) is now simplified as the following (real) process,

$$\begin{aligned} X(P) &= \sum_{m=-N}^{N} W_m(\phi)e^{im\lambda} = W_0(\phi) + \sum_{m=1}^{N} W_m(\phi)e^{im\lambda} + \sum_{m=-1}^{-N} W_m(\phi)e^{im\lambda} \\ &= W_0(\phi) + \sum_{m=1}^{N} W_m(\phi)e^{im\lambda} + \sum_{m=1}^{N} \overline{W_m(\phi)}e^{-im\lambda} \\ &= W_0(\phi) + \sum_{m=1}^{N} \big[(W_m^r(\phi) + iW_m^i(\phi))(\cos(m\lambda) + i\sin(m\lambda)) \\ &\quad + (W_m^r(\phi) - iW_m^i(\phi))(\cos(m\lambda) - i\sin(m\lambda))\big] \\ &= W_0(\phi) + 2\sum_{m=1}^{N} \big[W_m^r(\phi)\cos(m\lambda) - W_m^i(\phi)\sin(m\lambda)\big]. \qquad (5.1.5) \end{aligned}$$

5.1.1 Data generation

Now for each fixed $m = 0, 1, 2, \cdots, N$, we consider $W_m(\phi) = W_m^r(\phi) + iW_m^i(\phi)$ then $W_m^*(\phi) = W_m^r(\phi) - iW_m^i(\phi)$ (where $W_m^*(\phi)$ is the complex conjugate of $W_m(\phi)$). We may assume that $W_m^r(\phi)$ and $W_m^i(\phi)$ are independent, each following a (Gaussian) distribution with mean zero and the same variance $\sigma_m^2(\phi) = \frac{1}{2}C_m^r(\phi, \phi)$, ($C_m^i(\phi, \phi) = 0$ implies $W_m^r(\phi)$ and $W_m^i(\phi)$ are uncorrelated, or independent for Gaussian). In chapter 1 we introduced the concept of circularly-symmetry, thus according to Gallager (2008) a complex random variable is circularly-symmetric if and only if it pseudo covariance is zero (1.2.1). In this section we will show that the Gaussian random variable $W_m(\phi)$ is a circularly-symmetric complex random variable.

Now for a set of distinct latitudes $\Phi = \{\phi_1, \phi_2, \cdots, \phi_{n_l}\}$, we consider a sequence of complex random variables $\{W_m(\phi) : \phi \in \Phi\}$, which forms a multivariate complex random vector $\underset{\sim}{W}_m = (W_m(\phi_1), W_m(\phi_2), \cdots, W_m(\phi_n))^T$ where $W_m(\phi_i) = W_m^r(\phi_i) + iW_m^r(\phi_i)$ with associated $2 \times n_l$-dimensional real random vector

$$\underset{\sim}{V}_m = (W_m^r(\phi_1), W_m^i(\phi_1), W_m^r(\phi_2), W_m^i(\phi_2), \cdots, W_m^r(\phi_{n_l}), W_m^i(\phi_{n_l}))^T.$$

Now we calculate the covariance matrix $K_W = E(\underset{\sim}{W}_m \underset{\sim}{W}_m^*)$ (where $\underset{\sim}{W}_m^*$ is the conjugated transpose) and pseudo-covariance $M_W = E(\underset{\sim}{W}_m \underset{\sim}{W}_m^T)$. Further, from 1.2.1 a complex random vector is circularly-symmetric if and only if M_W is zero.

$$\begin{aligned}
M_W &= \begin{pmatrix} E[W_m(\phi_1)W_m(\phi_1)] & E[W_m(\phi_1)W_m(\phi_2)] & \cdots & E[W_m(\phi_1)W_m(\phi_{n_l})] \\ E[W_m(\phi_2)W_m(\phi_1)] & E[W_m(\phi_2)W_m(\phi_2)] & \cdots & E[W_m(\phi_2)W_m(\phi_{n_l})] \\ \vdots & \vdots & \ddots & \vdots \\ E[W_m(\phi_{n_l})W_m(\phi_1)] & E[W_m(\phi_{n_l})W_m(\phi_2)] & \cdots & E[W_m(\phi_{n_l})W_m(\phi_{n_l})] \end{pmatrix} \\
&= \mathbf{0}
\end{aligned}$$

We can show the above result for $\forall\ i, j$,

$$\begin{aligned}
& E[W_m(\phi_i)W_m(\phi_j)] \\
= \ & E[(W_m^r(\phi_i) + iW_m^i(\phi_i))(W_m^r(\phi_j) + iW_m^i(\phi_j))] \\
= \ & E(W_m^r(\phi_i)W_m^r(\phi_j)) - E(W_m^i(\phi_i)W_m^i(\phi_j)) + i[E(W_m^r(\phi_i)W_m^i(\phi_j)) + E(W_m^i(\phi_i)W_m^r(\phi_j))] \\
& \text{for } i \neq j \\
= \ & \frac{1}{2}(C_m^r(\phi_i, \phi_j) - C_m^r(\phi_i, \phi_j)) + i[-\frac{1}{2}C_m^i(\phi_i, \phi_j) + \frac{1}{2}C_m^i(\phi_i, \phi_j)] = 0 \\
& \text{for } i = j \\
= \ & \frac{1}{2}(C_m^r(\phi_i, \phi_i) - C_m^r(\phi_i, \phi_i)) + i[0 + 0] = 0 \quad ; W_m^r(\phi_i), W_m^i(\phi_i) \text{ are independent}
\end{aligned}$$

Therefore, $\underset{\sim}{W}_m$ is circularly-symmetric. In addition,

$$
\begin{aligned}
K_W &= E(\underset{\sim}{W}_m \underset{\sim}{W}_m^*) \\
&= \begin{pmatrix} E[W_m(\phi_1)W_m^*(\phi_1)] & E[W_m(\phi_1)W_m^*(\phi_2)] & \cdots & E[W_m(\phi_1)W_m^*(\phi_{n_l})] \\ E[W_m(\phi_2)W_m^*(\phi_1)] & E[W_m(\phi_2)W_m^*(\phi_2)] & \cdots & E[W_m(\phi_2)W_m^*(\phi_{n_l})] \\ \vdots & \vdots & \ddots & \vdots \\ E[W_m(\phi_{n_l})W_m^*(\phi_1)] & E[W_m(\phi_{n_l})W_m^*(\phi_2)] & \cdots & E[W_m(\phi_{n_l})W_m^*(\phi_{n_l})] \end{pmatrix} \\
&= \begin{pmatrix} C_m^r(\phi_1,\phi_1) & C_m^r(\phi_1,\phi_2)+iC_m^i(\phi_1,\phi_2) & \cdots & C_m^r(\phi_1,\phi_{n_l})+iC_m^i(\phi_1,\phi_{n_l}) \\ C_m^r(\phi_2,\phi_1)-iC_m^i(\phi_2,\phi_1) & C_m^r(\phi_2,\phi_2) & \cdots & C_m^r(\phi_2,\phi_{n_l})+iC_m^r(\phi_2,\phi_{n_l}) \\ \vdots & \vdots & \ddots & \vdots \\ C_m^r(\phi_{n_l},\phi_1)-iC_m^i(\phi_{n_l},\phi_1) & C_m^r(\phi_{n_l},\phi_2)-iC_m^r(\phi_{n_l},\phi_2) & \cdots & C_m^r(\phi_{n_l},\phi_{n_l}) \end{pmatrix} \\
&= \begin{pmatrix} C_m^r(\phi_1,\phi_1) & C_m^r(\phi_1,\phi_2) & \cdots & C_m^r(\phi_1,\phi_{n_l}) \\ C_m^r(\phi_2,\phi_1) & C_m^r(\phi_2,\phi_2) & \cdots & C_m^r(\phi_2,\phi_{n_l}) \\ \vdots & \vdots & \ddots & \vdots \\ C_m^r(\phi_{n_l},\phi_1) & C_m^r(\phi_{n_l},\phi_2) & \cdots & C_m^r(\phi_{n_l},\phi_{n_l}) \end{pmatrix} \\
&\quad +i \begin{pmatrix} 0 & C_m^i(\phi_1,\phi_2) & \cdots & C_m^i(\phi_1,\phi_{n_l}) \\ -C_m^i(\phi_2,\phi_1) & 0 & \cdots & C_m^i(\phi_2,\phi_{n_l}) \\ \vdots & \vdots & \ddots & \vdots \\ -C_m^i(\phi_{n_l},\phi_1) & -C_m^i(\phi_{n_l},\phi_2) & \cdots & 0 \end{pmatrix} \\
&= Re(K_W) + iIm(K_W)
\end{aligned}
$$

Now,

$$
K_V = E(\underset{\sim}{V}_m \underset{\sim}{V}_m^*) = E(\underset{\sim}{V}_m \underset{\sim}{V}_m^T)
$$

In order to generate K_V for n_l-tuple case, we reorganize the vector $\underset{\sim}{V}_m$ into the following form.

$$
\begin{aligned}
\underset{\sim}{V}_m &= (W_m^r(\phi_1),\cdots,W_m^r(\phi_{n_l}),W_m^i(\phi_1),\cdots,W_m^r(\phi_{n_l}))^T \\
&= (Re(\underset{\sim}{W}_m), Im(\underset{\sim}{W}_m))^T
\end{aligned}
$$

that is, we grouped all real components and imaginary components together. Hence,

$$
\begin{aligned}
K_V &= E(\underset{\sim}{V}_m \underset{\sim}{V}_m^T) \\
&= \begin{pmatrix} E[Re(\underset{\sim}{W}_m)Re(\underset{\sim}{W}_m)^T] & E[Re(\underset{\sim}{W}_m)Im(\underset{\sim}{W}_m)^T] \\ E[Im(\underset{\sim}{W}_m)Re(\underset{\sim}{W}_m)^T] & E[Im(\underset{\sim}{W}_m)Im(\underset{\sim}{W}_m)^T] \end{pmatrix}_{2n_l \times 2n_l}
\end{aligned}
$$

Since $\underset{\sim}{W}_m$ is circularly-symmetric from 1.2.3 we can get the following results,

$$E[Re(\underset{\sim}{W}_m)Re(\underset{\sim}{W}_m)^T] = E[Im(\underset{\sim}{W}_m)Im(\underset{\sim}{W}_m)^T] = \frac{1}{2}(Re(K_W))_{n_l \times n_l}$$
$$E[Re(\underset{\sim}{W}_m)Im(\underset{\sim}{W}_m)^T] = -E[Im(\underset{\sim}{W}_m)Re(\underset{\sim}{W}_m)^T] = \frac{1}{2}(Im(K_W))_{n_l \times n_l}$$

$$K_V \quad = \quad \frac{1}{2}\begin{pmatrix} Re(K_W) & Im(K_W)^T \\ Im(K_W) & Re(K_W) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} Re(K_W) & -Im(K_W) \\ Im(K_W) & Re(K_W) \end{pmatrix}$$

Since K_V is a non-negative definite and matrix, it can be represented as follows,

$$K_V = Q\Lambda Q^T,$$

where Λ is a diagonal matrix with eigen values (real-positive) of K_V and Q are the corresponding orthonormal eigenvectors. We can choose $A = Q\Lambda^{1/2}Q^T$ to obtain,

$$\underset{\sim}{V}_m = A_{2n_l \times 2n_l} Z_{2n_l \times 1},$$

where $Z = \{z_1, z_2, \ldots, z_{n_l}, z_1^*, z_2^*, \ldots, z_{n_l}^*\}$ and each $z_i \sim N(0,1)$ hence we can get $\underset{\sim}{W}_m$. Now for each latitude $\phi_l, l = 1, 2, \cdots, n_l$ and $\lambda_k, k = 1, 2, \cdots, n_L$ $(N = n_L/2)$, we denote the axially symmetric data (real) as $X(\phi_l, \lambda_k)$. These random variates can be obtained from the equation (5.1.5), let's rewrite the equation as follows,

$$X(\phi_l, \lambda_k) \quad = \quad W_0(\phi_l) + 2\sum_{m=1}^{N}\left[W_m^r(\phi_l)\cos(m\lambda_k) - W_m^i(\phi_l)\sin(m\lambda_k)\right] \qquad (5.1.6)$$

5.1.2 Pseudo-code

- Choose a cross covariance function, $R(P,Q)$
- Initialize the parameters (C_1, C_2, a, u, p) and choose a resolution $\phi_1, \ldots, \phi_{n_l}, \lambda_1, \ldots, \lambda_{n_L}$ (or $n_l \times n_L$),
- Derive $C_m(\phi_P, \phi_Q)$ based on $R(P,Q)$ where $m = 0, 1, \ldots, n_L/2$,
    1. for each m get $Re(K_W)$ and $Im(K_W)$ hence obtain K_V
    2. use SVD to get $\underset{\sim}{V}_m$ $(n_l - tuples)$
    3. get $\underset{\sim}{W}_m$'s from $\underset{\sim}{V}_m$
- apply the equation (5.1.6) to generate grid data.

5.1.3 Results

Simulated data sample:

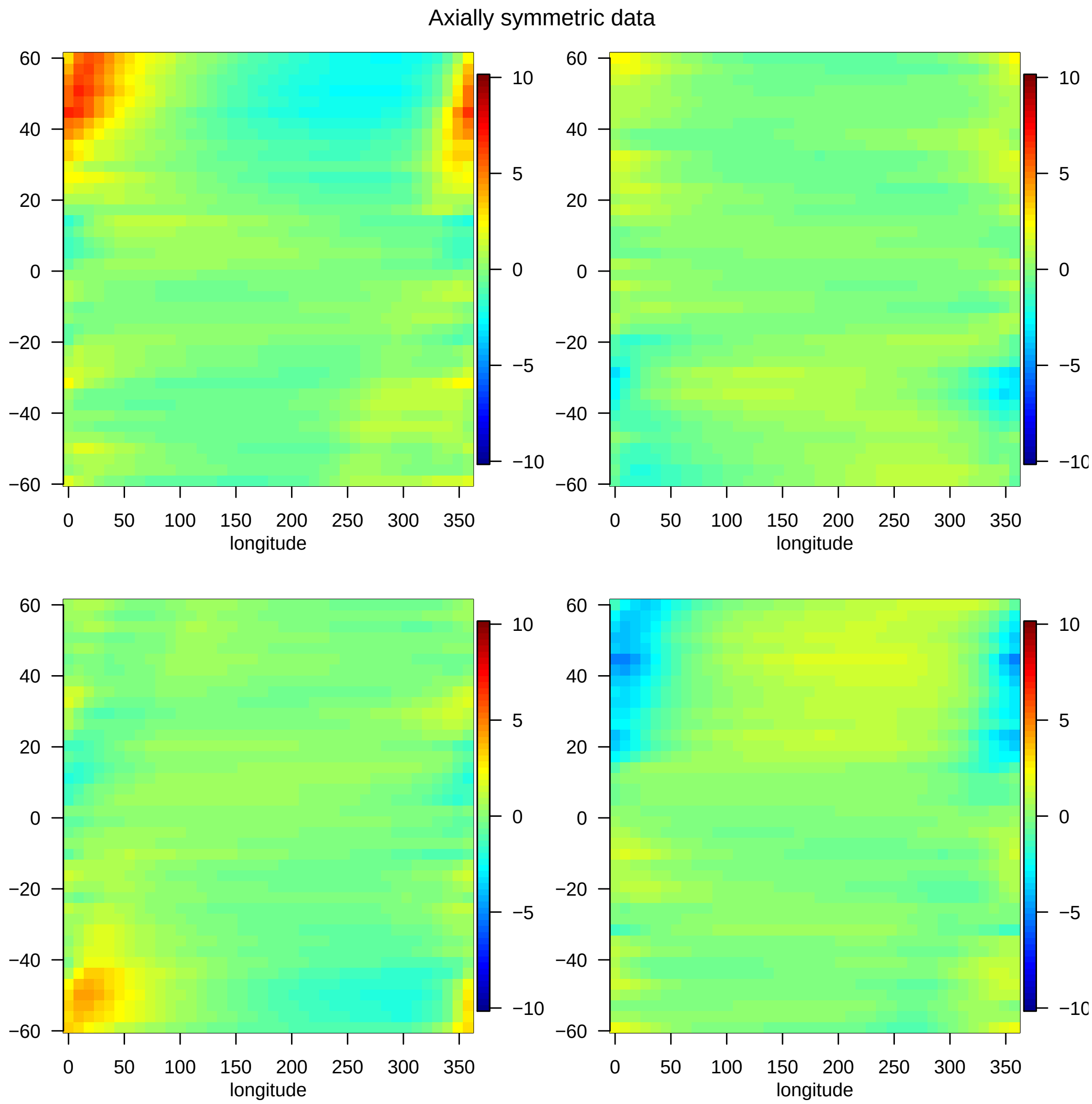

Figure 5.1: Four consecutive axially symmetric data snapshots based on model 2, grid resolution $2^0 \times 1^0$ (data scale -10 and 10).

Comprison of the proposed models with MOM estimates:

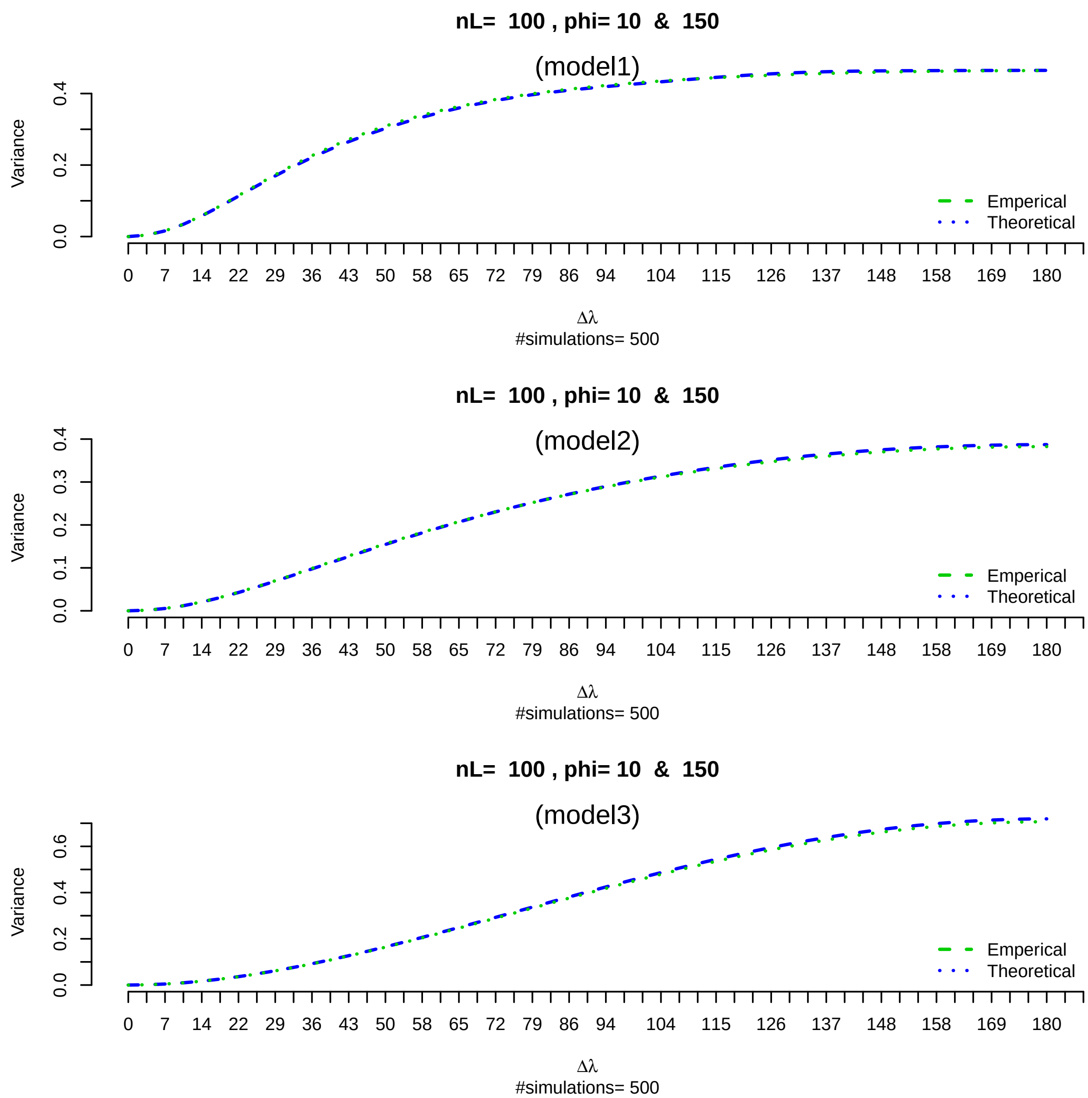

Figure 5.2: Variogram comparison when $u = 0$

- Model 1

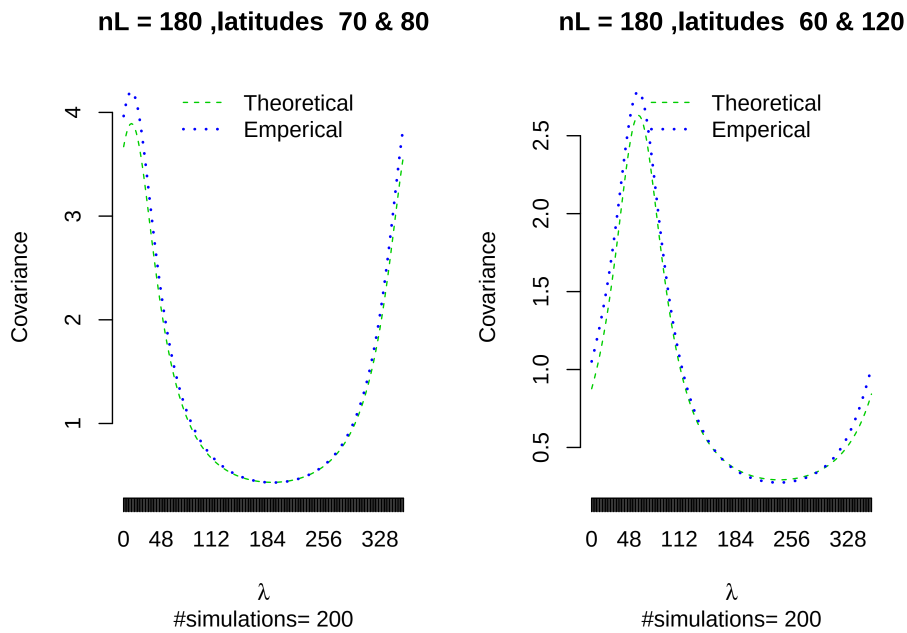

Figure 5.3: Cross covariance comparison of model1

- Model 2

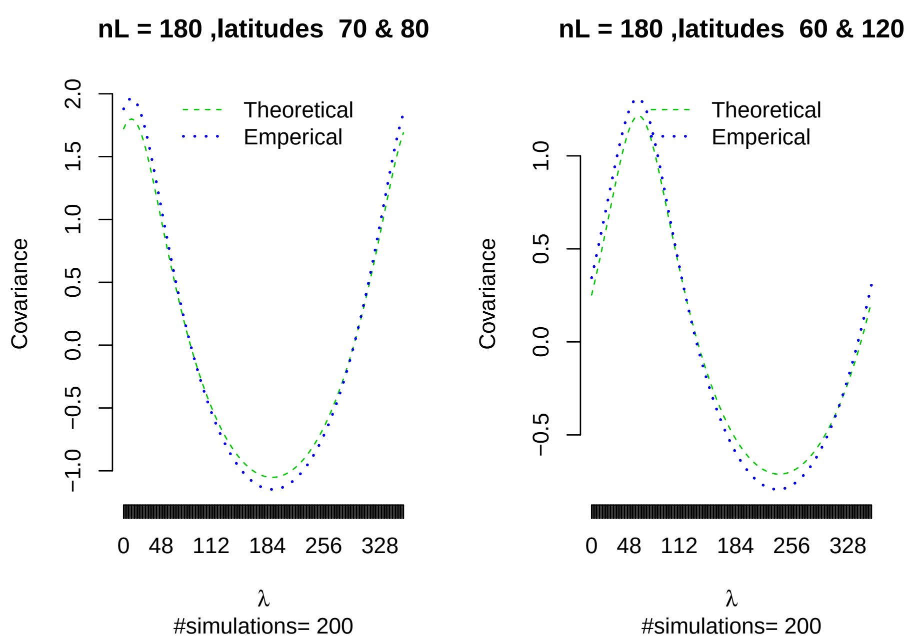

Figure 5.4: Cross covariance comparison of model2

- Model 3

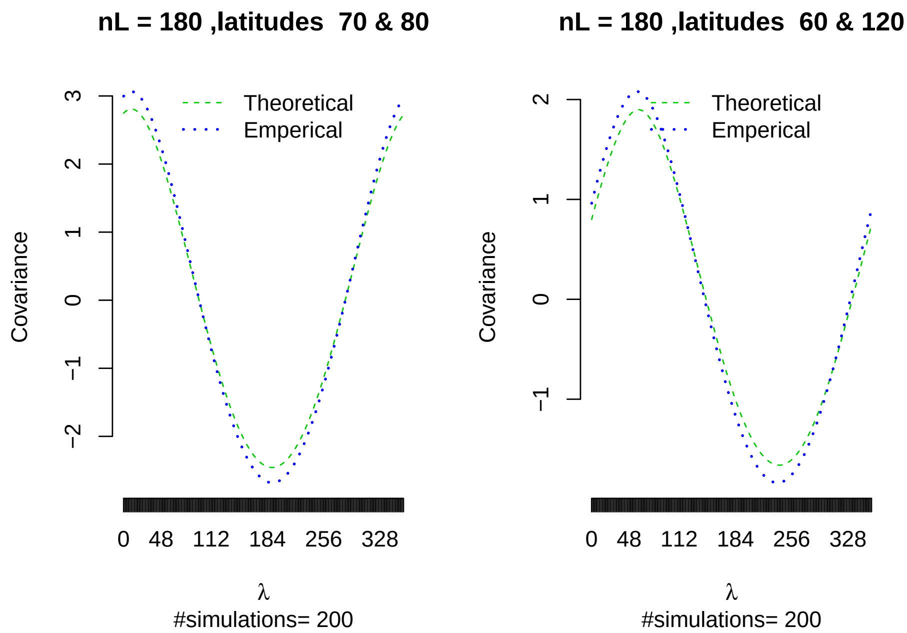

Figure 5.5: Cross covariance comparison of model3

Chapter 6

Future Research (due August 28)

Future research work !!

Bibliography

Bolin, D. and Lindgren, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *The Annals of Applied Statistics*, pages 523–550.

Christy, J. R., Spencer, R. W., and Braswell, W. D. (2000). Msu tropospheric temperatures: Dataset construction and radiosonde comparisons. *Journal of Atmospheric and Oceanic Technology*, 17(9):1153–1170.

Cressie, N. (1993). *Statistics for spatial data*. John Wiley & Sons.

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

De Mazancourt, T. and Gerlic, D. (1983). The inverse of a block-circulant matrix. *IEEE transactions on antennas and propagation*, 31(5):808–810.

Dufour, J.-M. and Roy, R. (1976). On spectral estimation for a homogeneous random process on the circle. *Stochastic Processes and their Applications*, 4(2):107 – 120.

Gallager, R. G. (2008). Circularly-symmetric gaussian random vectors. *preprint*, pages 1–9.

Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, 19(4):1327–1349.

Hitczenko, M. and Stein, M. L. (2012). Some theory for anisotropic processes on the sphere. *Statistical Methodology*, 9(1):211–227.

Huang, C., Zhang, H., and Robeson, S. M. (2011). On the validity of commonly used covariance and variogram functions on the sphere. *Mathematical Geosciences*, 43(6):721–733.

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

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

Jones, R. H. (1963). Stochastic processes on a sphere. *The Annals of mathematical statistics*, 34(1):213–218.

Jun, M. and Stein, M. L. (2008). Nonstationary covariance models for global data. *The Annals of Applied Statistics*, pages 1271–1289.

Katzfuss, M. and Cressie, N. (2011). Spatio-temporal smoothing and em estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, 32(4):430–446.

Leadbetter, H. C.-M. (1967). Stationary and related stochastic processes. *New York*.

Li, T.-H. and North, G. R. (1997). Aliasing effects and sampling theorems of spherical random fields when sampled on a finite grid. *Annals of the Institute of Statistical Mathematics*, 49(2):341–354.

Li, Y. (2013). Non-parametric and semi-parametric estimation of spatial covariance function.

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

Muir, S. T. (1920). *The theory of determinants*.

Roy, R. (1972). Spectral analysis for a random process on the circle. *Journal of Applied Probability*, pages 745–757.

Schoenberg, I. J. (1942). Positive definite functions on spheres. *Duke Math. J.*, 9(1):96–108.

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

Stein, M. L. (2007). Spatial variation of total column ozone on a global scale. *The Annals of Applied Statistics*, pages 191–210.

Tee, G. J. (2005). Eigenvectors of block circulant and alternating circulant matrices.

Trapp, G. E. (1973). Inverses of circulant matrices and block circulant matrices. *Kyungpook Mathematical Journal*, 13(1):11–20.

Wackernagel, H. (2013). *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media.

Yadrenko, M. I. (1983). *Spectral theory of random fields*. Optimization Software.

Yaglom, A. (1987). *Correlation Theory of Stationary and Related Random Functions: Vol.: 1: Basic Results*. Springer-Verlag.

Yaglom, A. M. (1961). Second-order homogeneous random fields. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob*, volume 2, pages 593–622.