

VANLANGENBERG, CHRISTOPHER D., Ph.D. Data Generation and Estimation for Axially Symmetric Processes on the Sphere . (2016)
Directed by Dr. Haimeng Zhang. 19 pp.

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.

DATA GENERATION AND ESTIMATION FOR AXIALLY SYMMETRIC
PROCESSES ON THE SPHERE

by

Christopher D. Vanlangenberg

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

Greensboro
2016

Approved by

Committee Chair

© 2016 by Christopher D. Vanlangenberg

In memory of my fater, Donald Anthony Vanlangenberg

APPROVAL PAGE

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

Committee Chair _____
Haimeng Zhang

Committee Members _____
Sat Gupta

Scott Richter

Xiaoli Gao

Shan Suthaharan

Date of Acceptance by Committee

Date of Final Oral Examination

ACKNOWLEDGMENTS

Complete later.....!

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER	
I. INTRODUCTION (DUE AUGTH 4TH)	1
II. LITERATURE REVIEW (DUE AUGUST 28)	8
III. COVARIANCE AND VARIOGRAM ESTIMATION ON THE CRICLE (DUE JULY 31)	13
IV. RANDOM PROCESS ON A SPHERE (DUE AUGUST 21)	15
4.1. Axially Symmetric Processes	15
4.2. Covariance and Variogram Estimation	15
V. GLOBAL DATA GENERATION ON THE SPHERE (DUE AUGUST 21)	16
5.1. Data Generation	16
5.2. Cross Variogram Estimation and Data Verifi- cation)	16
VI. FUTURE RESEARCH (DUE AUGUST 28)	17
REFERENCES	18

LIST OF TABLES

	Page
Table 1. Commonly used covariance and variogram models	12

LIST OF FIGURES

	Page
Figure 1. Theoretical and empirical covariance comparison com- parison on a circle	14
Figure 2. Theoretical and empirical covariance comparison com- parison on a circle	14

CHAPTER I

INTRODUCTION (DUE AUGTH 4TH)

Strict stationarity: for all finite n , $x_1, \dots, x_n \in \mathbb{R}^d$, $h_1, \dots, h_n \in \mathbb{R}$ and $x \in \mathbb{R}^d$,

$$P\{Z(x_1 + x) \leq h_1, \dots, Z(x_n + x) \leq h_n\} = P\{Z(x_1) \leq h_1, \dots, Z(x_n) \leq h_n\} \quad (1.1)$$

Weak stationarity: If a random process $Z(x)$ has a finite second moment, it's mean function is constant and the covariance is a function of distance $C(h) = Cov(Z(x), Z(x + h))$ also referred as auto-covariance).

Note: Strictly stationary random fields with finite second moments is also weakly stationary.

Intrinsic stationarity: If the variance between two locations depend only on the distance of two locations. Weak stationarity implies intrinsic stationarity.

Isotropy: The invariance property of stationarity due to rotations and reflections.

If $Z(x)$ is a random field on \mathbb{R}^d it is strictly isotropic if the joint distributions are invariant under all rigid motions. *i.e.*, for any orthogonal $d \times d$ matrix H and any $x \in \mathbb{R}^d$

$$P\{Z(Hx_1 + x) \leq h_1, \dots, Z(Hx_n + x) \leq h_n\} = P\{Z(x_1) \leq h_1, \dots, Z(x_n) \leq h_n\} \quad (1.2)$$

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

Autocovariance: Suppose $Z(x)$ is weakly stationary on \mathbb{R}^d with autocovariance function $C(h) = \text{cov}(Z(x), Z(x+h))$ then,

$$C(0) \geq 0$$

$$C(h) = C(-h)$$

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

Note: $C(\cdot)$ function is p.d. if

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

for any integer N , any constants a_1, a_2, \dots, a_N , and any locations x_1, x_2, \dots, x_N .

Some properties of a p.d. function

- (1) If C_1, C_2 are p.d. then $a_1 C_1 + a_2 C_2$ is p.d. $\forall a_1, a_2 \geq 0$
- (2) If C_1, C_2, \dots are p.d. and $\lim_{n \rightarrow \infty} C_n(x) = C(x)$, $\forall x \in \mathbb{R}^d$, then C is p.d.
- (3) If C_1, C_2 are p.d. then $C(x) = C_1(x)C_2(x)$ is p.d.

Semivariogram: an alternative to the covariance function proposed by Matheron.

For an intrinsically stationary random field $Z(s)$,

$$\begin{aligned} E[Z(s)] &= \mu, \\ \gamma(h) &= \frac{1}{2} \text{Var}(Z(s+h) - Z(s)), \end{aligned} \tag{1.3}$$

Where γ is the semivariogram and $\gamma(h) = C(0) - C(h)$ for a weakly stationary process with covariance function $C(h)$.

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, \dots and a random variable X defined on a common probability space. Define, $X_n \xrightarrow{L^2} X$ if, $E(X^2) < \infty$ and $E(X_n - X)^2 \rightarrow 0$ as $n \rightarrow \infty$. We can 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,

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

If $Z(x)$ is weak 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 iff $C(\cdot)$ is continuous at the origin.

Spectral methods: Sometimes it is 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 weakly stationary so does $Z(x)$. The covariance function can be defined as,

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

for any complex constants c_1, \dots, c_n , and any locations x_1, x_2, \dots, x_n ,

$$\sum_{i,j=1}^n c_i \bar{c}_j C(x_i - x_j) \geq 0 \tag{1.4}$$

Spectral representation of a random field: Suppose $\omega_1, \dots, \omega_n \in \mathbb{R}^d$ and let Z_1, \dots, Z_n be mean zero complex random variables with $E(Z_i \bar{Z}_j) = 0, i \neq j$

j and $E|Z_i|^2 = f_i$. Then the random sum

$$Z(x) = \sum_{k=1}^n Z_k e^{i\omega_k^T x}. \quad (1.5)$$

Then $Z(x)$ given above 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,

$$C(x) = \int_{\mathbb{R}^d} e^{i\omega^T x} F(d\omega) \quad (1.6)$$

where F is the so-called spectral distribution. For more details one can refer to [Ste99][p. 24], Here is a more general result from Bochner.

Theorem 1.0.1 (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 iff it can be represented as above, where F is a positive measure.

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

$$f(\omega) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\omega^T x} C(x) dx \quad (1.7)$$

Septral densities: Rational Functions that are even, nonnegative 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).

Septral densities: Gaussian Model 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 $f(\omega) = \frac{1}{2\sqrt{\pi\alpha}} ce^{\frac{-\omega^2}{4\alpha}}$.

Septral densities: Matern 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 is

$$C(h) = \frac{\pi^{1/2}\phi}{2^{\nu-1}\Gamma(\nu+1/2)\alpha^{2\nu}}(\alpha|h|)^{\nu}Y_{\nu}(\alpha|h|), \quad (1.8)$$

where Y_{ν} is the modified Bessel function, the larger the ν smoother the Y . Further, Y will be m times square differntiable 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 $e^{-\alpha|h|} \cdot \text{polynomial}(|h|)$

$$\nu = 1/2 \quad : \quad C(h) = \pi\phi\alpha^{-1}e^{-\alpha|h|}$$

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

Matérn Variance: We can use the covariance function given by (1.8) to compute the Matérn variances at different latitudes. The spectral density of the rational form was fitted to residuals of TOMS data.

CHAPTER II

LITERATURE REVIEW (DUE AUGUST 28)

- (1) Axially symmetry, which means that a process is invariant to rotations about the Earth's axis. The idea was first proposed by [Jon63], 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. [Ste07] argued that Total Ozone Mapping Spectrometer (TOMS) data varies strongly with latitudes and homogeneous models are not suitable. Further, [CJ08], [JS08], [BL11] 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. [Ste07] 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, [Ste99] with 170 parameters to estimate and [CJ08] more than 396 parameters to estimate.
- (5) When modeling spatial data stationary models are less useful; [JS08] 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) [LRL11] 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. [BL11] (continuation of the work proposed in [LRL11]) 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 [KC11]. They used MOM parameter estimation method (similar approach in FRS). This work is also based on [CJ08] 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. [HZR12] 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 [HZR12].

- (10) [HS12] 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 [Li13]. They proposed convolution methods to generate random fields with a class of *Matérn*-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) *Matérn* 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. [JJ15] proposed *Matérn*-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
<i>Matérn</i>	$\frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\frac{h}{\phi})^\nu Y_\nu(\frac{h}{\phi})$	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 1. Commonly used covariance and variogram models

CHAPTER III
COVARIANCE AND VARIOGRAM ESTIMATION ON THE CRICLE (DUE
JULY 31)

discuss about data generation on a circle,

- *including circulant matrix*
- *why circulant matrix*
- *discuss covarince, biasness and the difficulties of estimation*
-

Using the covarince function $C_1 e^{-a|\Delta\lambda|}$ and estimate empirical covarince from MOM as follows,

$$C(\theta) = \frac{1}{n_L} \sum_{i=1}^{n_L} (X(a_i + \theta) \cdot X(a_i)) - (\overline{X(a)})^2 \quad (3.1)$$

covariance estimator without the mean term

$$C(\theta) = \frac{1}{n_L} \sum_{i=1}^{n_L} (X(a_i + \theta) \cdot X(a_i)) \quad (3.2)$$

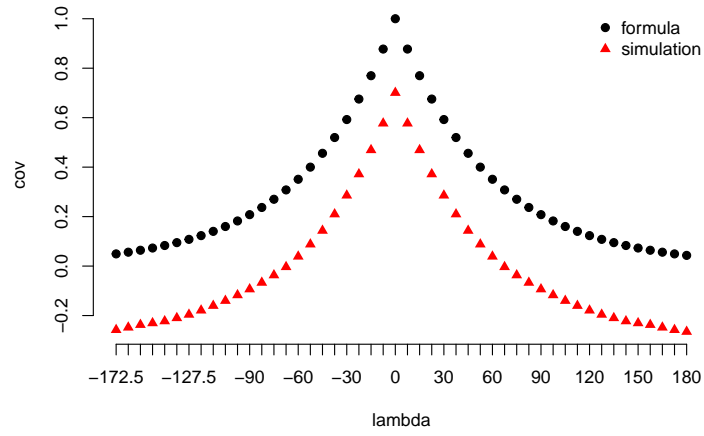


Figure 1. Theoretical and empirical covariance comparison comparison on a circle

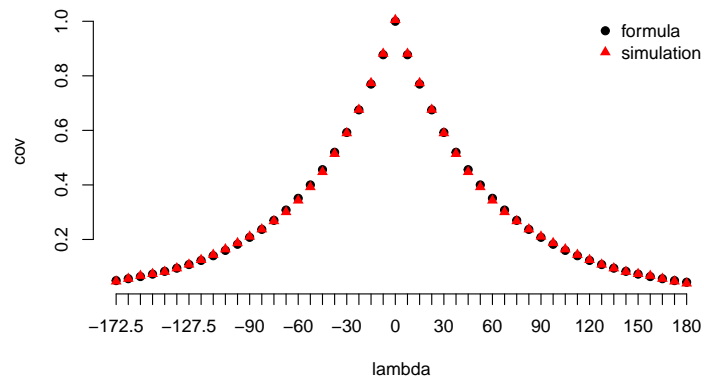


Figure 2. Theoretical and empirical covariance comparison comparison on a circle

CHAPTER IV

RANDOM PROCESS ON A SPHERE (DUE AUGUST 21)

4.1 Axially Symmetric Processes

4.2 Covariance and Variogram Estimation

CHAPTER V

GLOBAL DATA GENERATION ON THE SPHERE (DUE AUGUST 21)

5.1 Data Generation

5.2 Cross Variogram Estimation and Data Verification)

CHAPTER VI
FUTURE RESEARCH (DUE AUGUST 28)

Future research work !!

REFERENCES

- [BL11] David Bolin and Finn Lindgren, *Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping*, The Annals of Applied Statistics (2011), 523–550.
- [CJ08] Noel Cressie and Gardar Johannesson, *Fixed rank kriging for very large spatial data sets*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **70** (2008), no. 1, 209–226.
- [HS12] Marcin Hitczenko and Michael L Stein, *Some theory for anisotropic processes on the sphere*, Statistical Methodology **9** (2012), no. 1, 211–227.
- [HZR12] Chunfeng Huang, Haimeng Zhang, and Scott M Robeson, *A simplified representation of the covariance structure of axially symmetric processes on the sphere*, Statistics & Probability Letters **82** (2012), no. 7, 1346–1351.
- [JJ15] Jaehong Jeong and Mikyoung Jun, *A class of matérn-like covariance functions for smooth processes on a sphere*, Spatial Statistics **11** (2015), 1–18.
- [Jon63] Richard H Jones, *Stochastic processes on a sphere*, The Annals of mathematical statistics **34** (1963), no. 1, 213–218.
- [JS08] Mikyoung Jun and Michael L Stein, *Nonstationary covariance models for global data*, The Annals of Applied Statistics (2008), 1271–1289.
- [KC11] Matthias Katzfuss and Noel Cressie, *Spatio-temporal smoothing and em estimation for massive remote-sensing data sets*, Journal of Time Series Analysis **32** (2011), no. 4, 430–446.
- [Li13] Yang Li, *Non-parametric and semi-parametric estimation of spatial covariance function*.
- [LRL11] Finn Lindgren, Håvard Rue, and Johan Lindström, *An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **73** (2011), no. 4, 423–498.
- [Ste99] Michael L Stein, *Interpolation of spatial data: some theory for kriging*, Springer Science & Business Media, 1999.

- [Ste07] ———, *Spatial variation of total column ozone on a global scale*, The Annals of Applied Statistics (2007), 191–210.