Modeling Data Observed on Spheres and Graphs

By

AMY TAE YEN KIM

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICS

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Co-chair Debashis Paul, Ph.D.

Co-chair Thomas C.M. Lee, Ph.D.

Ethan Anderes, Ph.D.

Committee in Charge

2020

To Gina

Contents

Modeling Data observed on Spheres and Graphs

Abstract

This dissertation presents novel methods for two spatial modeling problems: tangential vector random fields on a sphere and stochastic processes on undirected graphs. The solutions use the specific characteristic of the spatial domains and provide new perspectives on characterizing the processes using the spectral decomposition of the Laplacian operator. The methods are used in various applications related to geophysics and economics.

Acknowledgments

I would like to express my sincere appreciation and deepest respect to my advisors, Professor Thomas C.M. Lee and Professor Debashis Paul. They share their invaluable experiences and useful critiques on my research and keep my progress on schedule. They have a willingness to give their time so generously. Thanks to their great support and patient guidance, I was able to complete my works. The values I have learned from their personalities and expertise in my Ph.D. years will guide and inspire me to become a better researcher. I am incredibly fortunate to have them as my advisors.

I would like to thank the members of my committees: Professor Ethan Anderes, Professor Jie Peng, and Professor Can Le for their encouragement, insightful comments, and challenging questions. I am also thankful to Professor Tomoko Matsuo and Professor Anandamayee Majumdar for working on the projects together. I gratefully acknowledge the financial support provided by the Department of Statistics at UC Davis.

I am grateful to Professor Peter Guttorp and Professor June Morita at the University of Washington, Seattle, for their guidance. They gave me the first opportunity of research experience, and they encouraged me to pursue a Ph.D.

I want to acknowledge the support and great love of my family, my husband, Chaechul; my parents-in-law Sunim and Guntaik; my parents Seokjoo and Kwango; and my child, Gina. They kept me standing all hardships, and this work would not have been achievable without their input.

CHAPTER 1

Overview

Spatial data structures are prevalent in many fields of study. Their covariance structures are commonly modeled based on Euclidean distances between locations that incorporate the notions of stationary and isotropy. Euclidean distances are not appropriate for capturing the characteristics of processes defined over non-Euclidean space, such as a sphere. In practice, the assumptions of stationarity and isotropy of physical processes are subject to empirical verification and physical plausibility. Moreover, models for covariances determined by the Euclidean distance are hard to extend to non-stationary or anisotropic fields. We propose classes of parametric models for Gaussian processes on a sphere and on an undirected graph, that utilize the geometric characteristics of the respective domains, though the spectral decomposition of the Laplacian operator. We study the properties of the models and develop efficient estimation and prediction procedures based on the likelihood method.

In Chapter 2, we propose a parametric model for tangential vector random fields (TVRFs) on the surface of a sphere based on a vector spherical harmonics (VSH) representation. We construct vector random fields on a sphere by making use of the spectral representations in the VSH basis with random coefficients. This representation of the vector field in terms of the VSH basis naturally incorporates the Helmholtz-Hodge decomposition, enabling us to decompose vector fields on a sphere uniquely as a sum of curl-free and divergence-free fields while satisfying natural physical constraints. We propose a likelihood-based method to estimate the model parameters, and this random effects representation facilitates efficient computation of the maximum likelihood estimates of the parameters. The model enables us to use a parametric bootstrap method for uncertainty quantification. When the data are on an equiangular grid, the computational efficiency can be enhanced by using a discrete VSH transformation. We conduct extensive numerical studies to illustrate the estimation, model selection, and prediction performance of the proposed method. We

apply the proposed methodology to analyze the Orsted satellite-based measurements on the Earth's magnetic field.

In Chapter 3, we consider a class of stochastic processes observed on a set of nodes such that the processes are characterized using the graph Laplacian associated with an undirected graph on these nodes. These processes generalize stationary processes from temporal or spatial domains to the setting where the domain is an undirected graph. We construct the model using spectral decomposition of the graph Laplacian and consider several parametric models for such processes, assuming that the graph structure is known from extraneous information. We study inference questions on geometric graphs for the associated parameters. We examine the effects of graph sizes, missing nodes, and node structures on our model estimation and prediction performance. As an illustration, we apply this method to study the per capita gross domestic product (GDPc) across several counties, where the associated graph is the neighborhood graph.

CHAPTER 2

Tangential Vector Random Fields on a Sphere Using VSH

2.1. Introduction

Modeling vector random fields (VRFs) on a sphere has received considerable attention in recent years in many applications. In geophysics, Earth's magnetic field is a well-known example of a VRF on a sphere. In the atmospheric science, surface velocity winds and ocean currents are modeled as tangential VRFs (TVRFs) on Earth. Researchers also study the Helmholtz-Hodge decomposition of vector fields while retaining their geometric features. The Helmholtz-Hodge theorem implies that any vector fields on a sphere can be decomposed uniquely into a divergence-free component and a curl-free component. For instance, the divergence-free part of the horizontal wind field offers details about cyclonic storms, whereas the curl-free field offers features on high and low pressures systems [22][23]. Watterson [47] explains that the gyres within each ocean basin are divergence-free fields, and the overturning flows are in the irrotational field. Therefore, the decomposition can provide useful diagnostic information in associated applications [14]. Thus, it is essential to preserve the geometric features of the vector field during the estimation and predictions in modeling.

Applicable approaches to modeling VRFs on a sphere include treating the vector fields as multivariate random processes. Gneiting et al. [15] introduce the bivariate isotropic multivariate Matérn model and Apanasovich et al. [1] develop the multivariate stationary Matérn model with any number of components, which produces flexible parameter selections in each component. Kleiber and Nychka [28] develop a parametric non-stationary multivariate model with locally-varying parameters. It is not straightforward to extend these models to a sphere, as a proper distance metric is not readily available.

Other approaches to modeling multivariate random fields on a sphere include the use of differential operators. Jun and Stein [27] focus on the construction and characterization of covariance functions for Gaussian random fields on a sphere. Stein [46] models axially-symmetric processes

on a sphere using covariance functions in the form of scalar spherical harmonics (SSH). Jun [25] applies differential operators to multivariate isotropic spatial processes to construct non-stationary processes on a globe. Jun [26] proposes Matérn based non-stationary cross-covariance models for processes in a sphere with flexible spatially varying cross-correlation coefficients, which are smoothness parameters coupled with the differential operators. Those models do not incorporate any natural physical constraints on the VRF, such as the fields being divergence-free, as is often encountered in geophysical applications.

Narcowich et al. [38] present a method of constructing a divergence-free tangential vector field to surfaces in $\mathbb{R}^3$ with radial basis functions (RBFs). It uses the Hessian matrix of the RBFs in Cartesian coordinates, and the output must be converted to spherical coordinates. Fuselier and Wright [14] construct a tangential vector field on a sphere with RBFs. They implement surface-gradient and surface-curl on the RBFs, and the constructed tangential vector field can be decomposed into divergence-free and curl-free fields. Their construction includes the physical constraints, but they do not manifest on TVRFs.

Recently, Fan et al. [10] propose the Tangent Matérn model, which constructs a tangential random field on a unit sphere that contains the physical constraints. They apply a surface gradient and surface-curl operators to a scalar isotropic potential field to construct random fields. Its cross-covariance structure is chosen as a bivariate Matérn Model.

We propose a parametric model for VRFs on the surface of a sphere using VSH representations. The model provides a unified framework for modeling both mean and residuals fields within the linear mixed effects model framework. Because of the construction of the VSH basis itself, the resulting process naturally admits a Helmholtz-Hodge decomposition whose components are modeled parametrically. Thus, this model produces a physically interpretable decomposition of vector fields on a sphere and preserves the geometric features of the field even for prediction. The computational efficiency can be enhanced with an equiangular grid by using a discrete VSH transform. Moreover, the proposed model enables more accurate and principled uncertainty quantification of various quantities of interest using either a parametric bootstrap procedure or an asymptotic theory.

We illustrate the proposed VRF model by applying it to describe Earth's main magnetic field based on the Ørsted satellite survey data. In geophysics, modeling the main magnetic fields has

been actively researched due to its importance and challenges. The typical approach to modeling the geomagnetic vector field is to model it as the gradient of a scalar potential field, while the latter is modeled deterministically using a scalar spherical harmonics (SSH) representation [24]. CHAOS 6 [11] is currently one of the most accurate main magnetic field models and is likewise derived from the scalar potential field modelings. We apply the proposed model to the geomagnetic vector field, and the results show the benefit of modeling the random fluctuations over competing deterministic models in terms of improved fit and prediction accuracy. We also provide a principled approach to uncertainty quantification for the fitted vector field, which is an important contribution in this application domain.

The remainder of this article is organized as follows. In Section 2.2, we discuss the construction of VSH from SSH and connect it to Helmholtz-Hodge decomposition. We construct a tangential vector random field (TVRF) on the unit sphere through its representation in the VSH basis and propose a parametric Gaussian model with its covariance structure in Section 2.3. In Section 2.4, we discuss the maximum likelihood estimators (MLEs) of our model and best linear unbiased predictors (BLUPs). We conduct simulation studies in Section 2.5 to explore the empirical performance of our model, and we illustrate our model for application to geomagnetic fields in Section 2.6.

2.2. Vector Spherical Harmonics

Vector spherical Harmonics (VSH) have a fundamental role in our construction of VRFs on a sphere, so we review its definition and properties.

2.2.1. SSH and VSH. The complex scalar spherical harmonic, $Y_{l,m}(\theta,\phi)$, of degree l and order m are defined as

$$Y_{l,m}(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_{l,m}(\cos\theta)\exp(im\phi), \qquad m \geq 0 \tag{2.1}$$

$$Y_{l,m}(\theta,\phi) = (-1)^m Y^*_{l,-m}(\theta,\phi), \qquad m < 0 \tag{2.2}$$

where $l = 0, 1, 2, ..., -l \leq m \leq l$, $0 \leq \theta < \pi, 0 \leq \phi < 2\pi$. θ is a polar angle (colatitude), and ϕ is an azimuthal angle in a spherical coordinate system. $\{P_{l,m}\}$ is the associated Legendre function with degree l and order m, where $l = 0, 1, 2, \cdots$ and $-l \leq m \leq l$. $Y^*_{l,m}(\theta,\phi)$ is the complex conjugate of

$Y_{l,m}(\theta, \phi)$. SSH satisfy Equation (2.3) on $\mathbb{S}^2 = \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\| = 1\}$.

$$\Delta Y_{l,m} = l(l+1)Y_{l,m}, \tag{2.3}$$

where Δ is a spherical Laplacian operator. This implies that SSH are eigenfunctions of the spherical Laplacian on $\mathbb{S}^2$. We define the L^2 inner product of two complex functions F and G defined on $\mathbb{S}^2$ to be

$$\langle F, G \rangle = \int_{\mathbb{S}^2} F(\theta, \phi) G^*(\theta, \phi) \sin\theta d\theta d\phi,$$

where g^* denotes the complex conjugate of g. It is well-known that the SSH are orthonormal basis functions of $L^2(\mathbb{S}^2)$, that is, the inner product of $Y_{l,m}$ and $Y_{l',m'}$ is 1 if $l = l', m = m'$; otherwise the value is 0. Therefore, any complex-valued functions, $T(\theta, \phi)$, in $L^2(\mathbb{S}^2)$ can be expressed as a linear combination of scalar spherical harmonics as

$$T(\theta, \phi) = \sum_{l,m} f_{l,m} Y_{l,m}(\theta, \phi), \tag{2.4}$$

where

$$f_{l,m} = \int_{\mathbb{S}^2} T(\theta, \phi) Y^*_{l,m}(\theta, \phi) \sin\theta d\theta d\phi \tag{2.5}$$

and convergence holds in the $L^2(\mathbb{S}^2)$ sense. (*Peter-Weyl Theorem for a sphere* [34]). The complex coefficient $f_{l,m}$ is referred to as a scalar spherical transform of the function T, and the coefficient is calculated using the inner product of the function and SSH, $\langle T, Y_{l,m} \rangle$, in Equation (2.5).

VSH are constructed from SSH by application of appropriate differential operators [18]. The complex VSH functions defined on an annular shell consist of the triplets $\{\mathbf{Y}_{l,m}(\theta, \phi), \mathbf{B}_{l,m}(r, \theta, \phi), \mathbf{C}_{l,m}(r, \theta, \phi)\}$ where (r, θ, ϕ) is a point in a spherical shell. We express the basis elements of the VSH basis in the spherical coordinate system by making use of the canonical basis vectors, namely, $(\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$, where $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$ denote the canonical orthonormal basis vectors of the tangent space, and $\hat{\boldsymbol{r}}$ denotes the unit normal vector, respectively, on the surface of the r-sphere $\mathbb{S}^2(r) = \{x \in \mathbb{R}^3 :$

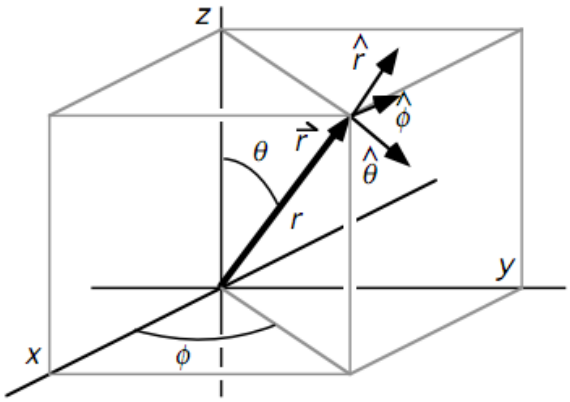


FIGURE 2.1. Spherical coordinates system

$\|x\|^2 = r^2\}$.

$$\mathbf{Y}_{l,m}(r,\theta,\phi) = Y_{l,m}(\theta,\phi)\hat{\mathbf{r}}, \tag{2.6}$$

$$\mathbf{B}_{l,m}(r,\theta,\phi) = r\nabla Y_{l,m}(\theta,\phi), \tag{2.7}$$

$$\mathbf{C}_{l,m}(r,\theta,\phi) = \vec{r}\times\nabla Y_{l,m}(\theta,\phi) = -\hat{\mathbf{r}}\times\mathbf{B}_{l,m}, \tag{2.8}$$

where $\vec{r}$ is the position vector of the point with spherical coordinates. $\mathbf{Y}_{l,m}$ are therefore seen as forming a basis for the radial component of a vector field defined on an annular shell. $\mathbf{B}_{l,m}$ and $\mathbf{C}_{l,m}$ are functions that form the basis of a vector field that is tangential to the r-sphere $\mathbb{S}^2(r)$.

The set of VSH functions are orthogonal to each other with respect to the inner product of two tangential vector fields, $\mathbf{v}$ and $\mathbf{u}$, defined as

$$\langle \mathbf{v}, \mathbf{u}\rangle = \int_{\mathbb{S}^2} (\mathbf{v}^\theta \mathbf{u}^{\theta *} + \mathbf{v}^\phi \mathbf{u}^{\phi *}) \sin\theta d\theta d\phi,$$

where any tangential vector field $\mathbf{v}$ is written with the tangential basis vectors as $\mathbf{v} = \mathbf{v}^\theta\hat{\boldsymbol{\theta}} + \mathbf{v}^\phi\hat{\boldsymbol{\phi}}$, and $\mathbf{u}^{\theta *}$ is a complex conjugate of $\mathbf{u}^\theta$. The definitions of VSH, (2.6, 2.7, 2.8), along with the relationships among the surface gradient and curl operators ensure orthogonality among the functions $\mathbf{Y}_{l,m}, \mathbf{B}_{l,m}, \mathbf{C}_{l,m}$ such as $\langle \mathbf{Y}_{l,m}, \mathbf{B}_{l,m}\rangle = 0, \langle \mathbf{Y}_{l,m}, \mathbf{C}_{l,m}\rangle = 0, \langle \mathbf{B}_{l,m}, \mathbf{C}_{l,m}\rangle = 0$ for the same degree l and order m. They are also orthogonal across pairs $(l,m) \neq (l',m')$.

The collection of tangential vector fields $\{\mathbf{B}_{l,m}, \mathbf{C}_{l,m}\}$ for $1 \leq l, |m| \leq l$ forms a complete basis of orthogonal functions in the space of all tangential vector fields defined on the r-sphere, for any fixed r, which henceforth, we set to be 1 unless otherwise specified.

The set of VSH $\{\mathbf{Y}_{l,m}, \mathbf{B}_{l,m}, \mathbf{C}_{l,m}\}$ has orthogonality and completeness as the SSH [2]. Any vector field $\mathbf{v}(r, \theta, \phi)$ on $\mathbb{S}^2$ is expanded in terms of VSH.

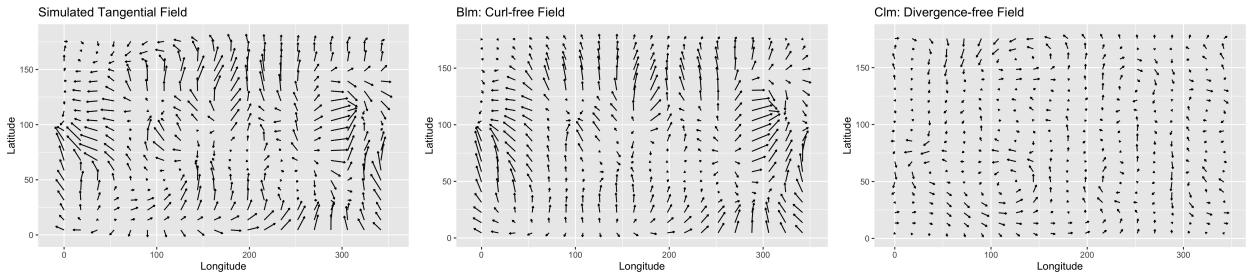
$$\mathbf{v}(r,\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left[f^R_{l,m}(r)\mathbf{Y}_{l,m}(r,\theta,\phi) + f^B_{l,m}(r)\mathbf{B}_{l,m}(r,\theta,\phi) + f^C_{l,m}(r)\mathbf{C}_{l,m}(r,\theta,\phi) \right], \tag{2.9}$$

where the complex coefficients $\{f^R_{l,m}(r), f^B_{l,m}(r), f^C_{l,m}(r)\}$ are referred to as vector spherical transforms. The coefficients are calculated as the inner products $f^R_{l,m}(r) = \langle \mathbf{v}, \mathbf{Y}_{l,m} \rangle$, $f^B_{l,m}(r) = \langle \mathbf{v}, \mathbf{B}_{l,m} \rangle$, and $f^C_{l,m}(r) = \langle \mathbf{v}, \mathbf{C}_{l,m} \rangle$.

We connect the linear combinations of VSH on a sphere to the Helmholtz-Hodge decomposition [13]. The decomposition states that any vector field $\mathbf{v}$ is uniquely decomposed as a sum of a divergence-free field, a curl-free field, and a harmonic component. The harmonic component vanishes on a closed space like on $\mathbb{S}^2$. The Helmholtz-Hodge decomposition implies that any tangential vector field $\mathbf{v}$ on $\mathbb{S}^2$ is uniquely decomposed [14] as a sum of a curl-free field and a divergence-free field. Specifically, a tangential vector field $\mathbf{v}$ defined on the unit sphere ($r = 1$) is expressed as a linear combination of $\mathbf{B}_{l,m}$ and $\mathbf{C}_{l,m}$ such that

$$\mathbf{v}(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \Big[\underbrace{f^B_{l,m}\mathbf{B}_{l,m}(\theta,\phi)}_{\text{Curl-free}} + \underbrace{f^C_{l,m}\mathbf{C}_{l,m}(\theta,\phi)}_{\text{Divergence-free}} \Big]. \tag{2.10}$$

The curl of $\mathbf{B}_{l,m}$ and the divergence of $\mathbf{C}_{l,m}$ are zero on $\mathbb{S}^2$. The tangential vector field $\mathbf{v}$ is the linear combination of the functions $\mathbf{B}_{l,m}$ represents the curl-free component of $\mathbf{v}$, and that in terms of $\mathbf{C}_{l,m}$ represents the divergence-free component of $\mathbf{v}$ (2.10). Thus, the VSH basis naturally yields the Helmholtz-Hodge decomposition of a tangential vector field on a sphere. This means, such a representation can be leveraged to model physical processes, such as geomagnetism, that are known to admit physical constraints such as being curl-free or divergence-free. The tangential vector fields with its curl-free field and divergence-free field are illustrated in Figure 2.2.



(a) Tangential Vector Field = (b) Curl-Free Field + (c) Divergence-Free Field

FIGURE 2.2. Simulated Tangential Vector Field: 20x20 Equiangular grid, $\sigma_B^2 = 20, \sigma_C^2 = 10, \alpha_B = 2, \alpha_C = 2, L = 9$

2.2.2. Implementation of VSH. VSH are obtained from evaluations of SSH using a recurrence relationship. Kostelec et al. [29] describe how one can use the evaluations of SSH functions to compute the vector spherical harmonic transforms. Our implementations of VSH basis computation and the VSH transform of a tangential vector field are based on this work.

The recurrence relationship of associated Legendre polynomials yields the following equations.

$$\mathbf{B}_{l,m}(\theta,\phi) = \frac{1}{\sin\theta}\Big[\big(c_{l,m}^{(1)} Y_{l+1,m}(\theta,\phi) - c_{l,m}^{(2)} Y_{l-1,m}(\theta,\phi)\big)\hat{\boldsymbol{\theta}} + ic_{l,m}^{(3)} Y_{l,m}(\theta,\phi)\hat{\boldsymbol{\phi}}\Big], \tag{2.11}$$

$$\mathbf{C}_{l,m}(\theta,\phi) = \frac{1}{\sin\theta}\Big[ic_{l,m}^{(3)} Y_{l,m}(\theta,\phi)\hat{\boldsymbol{\theta}} - \big(c_{l,m}^{(1)} Y_{l+1,m}(\theta,\phi) - c_{l,m}^{(2)} Y_{l-1,m}(\theta,\phi)\big)\hat{\boldsymbol{\phi}}\Big], \tag{2.12}$$

where $c_{l,m}$ are normalizing constants with given l, m such that

$$c_{l,m}^{(1)} = \frac{1}{\sqrt{l(l+1)}} \frac{l(l-m+1)}{(2l+1)} \sqrt{\frac{(2l+1)}{(2l+3)} \frac{(l+m+1)}{(l-m+1)}}, \tag{2.13}$$

$$c_{l,m}^{(2)} = \frac{1}{\sqrt{l(l+1)}} \frac{(l+1)(l+m)}{(2l+1)} \sqrt{\frac{(2l+1)}{(2l-1)} \frac{(l-m)}{(l+m)}}, \tag{2.14}$$

$$c_{l,m}^{(3)} = \frac{m}{\sqrt{l(l+1)}}. \tag{2.15}$$

VSH transforms can also be correspondingly derived from SSH transforms.

$$f_{l,m}^{B} = c_{l,m}^{(1)} g_{l+1,m}^{\theta} - c_{l,m}^{(2)} g_{l-1,m}^{\theta} - ic_{l,m}^{(3)} g_{l,m}^{\phi} \tag{2.16}$$

$$f_{l,m}^{C} = -ic_{l,m}^{(3)} g_{l,m}^{\theta} - c_{l,m}^{(1)} g_{l+1,m}^{\phi} + c_{l,m}^{(2)} g_{l-1,m}^{\phi} \tag{2.17}$$

where $g^{\theta}_{l,m}, g^{\phi}_{l,m}$ are SSH transforms that can be calculated based on the inner product with each components of the tangential vector field. $g^{\theta}_{l,m} = \left\langle \frac{1}{\sin\theta}\mathbf{v}^{\theta}, Y_{l,m} \right\rangle$,$g^{\phi}_{l,m} = \left\langle \frac{1}{\sin\theta}\mathbf{v}^{\phi}, Y_{l,m} \right\rangle$. After we obtain the scalar spherical transforms $\{g^{\theta}_{l,m}, g^{\phi}_{l,m}\}$, we can calculate the VSH transforms $\{f^{B}_{l,m}, f^{C}_{l,m}\}$. Notice the recurrence relations (2.16) and (2.17) show that in order to compute the VSH coefficients of $\mathbf{v}$, one needs to compute the SSH coefficients of the coordinate fields $\mathbf{v}^{\theta}$ and $\mathbf{v}^{\phi}$ up to an additional harmonic order beyond the maximum harmonic order of the VSH coefficients.

2.2.3. Evaluations of VSHs on an Equiangular Grid. The computations of VSH coefficients can be enhanced by fast Fourier transforms on $\mathbb{S}^2$. Driscoll and Healy [8] developed an effective algorithm for computing Fourier transforms on $\mathbb{S}^2$, which is a fast approach to computing SSH transforms. Healy Jr. et al. [17] reformulated their algorithm to improve inverse transforms and convolutions. The function, $T(\theta, \phi) \in L^2(\mathbb{S}^2)$ (2.4) is said to be band-limited with bandwidth $B \geq 0$ if $f_{l,m} = 0$ for all $l \geq B$, where $f_{l,m} = \langle T, Y_{l,m} \rangle$. For the band-limited functions, an equiangular grid sampling reduces the integrals (2.5) to finite weighted sums of the sampled data using the *Shannon's Sampling Theorem* (Theorem 1 [17]), so the integral can be computed using a discrete weighted sum. Kostelec and Rockmore [30] propose `S2kit`, a `C` routine, to compute discrete SSH transforms and their inverse transforms. We manipulate discrete VSH transforms and their inverse transform functions in `MATLAB` based on the `S2kit` package by Kostelec and Rockmore [30]. We determine an equiangular grid based on bandwidth B, which corresponds with the maximum degree of VSH. With bandwidth B, the $2B \times 2B$ equiangular grid on $\mathbb{S}^2$ has $(2B)^2$ points, and the point (θ_j, ϕ_k) are defined by $\theta_j = \pi(2j+1)/4B, \phi_k = 2\pi k/2B$, where $j = 0, 1, \ldots, 2B-1$ and $k = 0, 1, \ldots, 2B-1$. We extend the algorithm to compute VSH transforms and inverse VSH transforms based on the relationships (2.11, 2.12, 2.16, 2.17).

2.3. Modeling VRFs on a Sphere

2.3.1. TVRF Model. We introduce a statistical model of VRFs on a sphere by using VSH. Since SSH form a complete orthonormal basis for $L^2(\mathbb{S}^2)$, scalar random fields on a sphere can be represented by linear combinations of SSH. Marinucci and Peccati [34] present a Gaussian isotropic random field on a sphere based on its spectral representation in the SSH basis. The Gaussian

isotropic random field, $T = \{T(x) : s = (\theta, \phi) \in \mathbb{S}^2\}$, has the following spectral representation:

$$T(s) = \sum_{l=0}^{L} T_l(s) = \sum_{l=0}^{L} \sum_{m=-l}^{l} f_{l,m} Y_{l,m}(s), \tag{2.18}$$

where $T_l(s)$ is the orthogonal projection of the process T_l in the SH subspace associated with the l-th harmonic frequency, i.e., the subspace generated by $\{Y_{l,m} : -l \leq m \leq l\}$, for each fixed $l \geq 0$. The SSH coefficients $f_{l,m}$ are complex-valued Gaussian random variables and independent across both l and m (subject to conjugacy relationship for real-valued T). The Gaussian isotropic random field is a sum of L (it could be ∞) independent random processes.

The spectral representation (2.18) describes the scalar process $T(\cdot)$ on $\mathbb{S}^2$ as a random effects model. We extend this idea to construct a class of Gaussian *tangential vector random field (TVRFs)* on $\mathbb{S}^2$ by utilizing the VSH basis introduced in Section 2.2. Accordingly, a TVRF on $\mathbb{S}^2$ is expressed as a linear combination of the tangential VSH basis functions, $\{\mathbf{B}_{l,m}, \mathbf{C}_{l,m}\}$, with random coefficients.

The proposed class of TVRF, referred to as Gaussian isotropic TVRF, $\{X(s) : s \in \mathbb{S}^2\}$, is a tangential process on $\mathbb{S}^2$, with real coordinates, defined as

$$X(s) = \sum_{l=1}^{L} \sum_{m=-l}^{l} \Big[f^B_{l,m} \mathbf{B}_{l,m}(s) + f^C_{l,m} \mathbf{C}_{l,m}(s) \Big], \tag{2.19}$$

where the VSH coefficients, $\{f^B_{l,m}, f^C_{l,m}\}$, are complex Gaussian random variables and independent across l. L is the highest degree of VSH.

The VSH coefficients have complex Gaussian distributions with a mean of 0 and variances $\sigma^2_{B,l}, \sigma^2_{C,l}$. We denote $f^B_{l,m} \sim CN(0, \sigma^2_{B,l}), f^C_{l,m} \sim CN(0, \sigma^2_{C,l})$, which indicates

$$\begin{cases} \Re(f^\dagger_{l,m}) \sim N(0, \frac{1}{2}\sigma^2_{\dagger,l}) & m > 0 \\ \Re(f^\dagger_{l,m}) \sim N(0, \sigma^2_{\dagger,l}) & m = 0 \\ \Re(f^\dagger_{l,m}) = (-1)^{|m|} \Re(f^\dagger_{l,-m}) & m < 0 \end{cases}, \quad \begin{cases} \Im(f^\dagger_{l,m}) \sim N(0, \frac{1}{2}\sigma^2_{\dagger,l}) & m > 0 \\ \Im(f^\dagger_{l,m}) = 0 & m = 0 \\ \Im(f^\dagger_{l,m}) = (-1)^{|m|+1} \Im(f^\dagger_{l,-m}) & m < 0 \end{cases},$$

where $\Re(f^\dagger_{l,m})$ is a real part of $f^\dagger_{l,m}$, and $\Im(f^\dagger_{l,m})$ is an imaginary part, $\dagger \in \{B, C\}$. The coefficients satisfy the complex conjugate relationship, $f^\dagger_{l,m} = (-1)^m \{f^\dagger_{l,m}\}^*$.

The random coefficients are independent across l, and for a fixed l, $\Re(f^{\dagger}_{l,m})$ and $\Im(f^{\dagger}_{l,m})$, are iid across for all non-negative m. We introduce smoothness parameters, α_B, α_C for each field, such that the variances of the random coefficients decrease as the degree (l) of VSH increases.

$$\sigma^2_{\dagger,l} = (\frac{1}{l})^{\alpha_\dagger}\sigma^2_{\dagger}, \qquad \dagger \in \{B, C\} \tag{2.20}$$

The decay parameter is assumed to be greater than or equal to 1. When $L = \infty$, we set $\alpha \geq 1$ to ensure a finite variance. With a finite L, α can be less than 1 and greater than 0. The algebraic decays (2.20) indicate that the TVRFs can be formulated with finite L basis functions.

We assume that we observe bivariate data of the form $y(s_i) = X(s_i)+\epsilon_i$, where the observational noise ϵ_i is assumed to be i.i.d. $N(0, \tau^2 I_2)$, with I_2 denoting the 2×2 identity matrix. Therefore, by using the conjugacy relationship among the coefficients of the TVRF $X(s)$, we can express the observations $\{y(s_i)\}_{i=1}^n$, in terms of real and imaginary parts of the bases $\{\mathbf{B}_{l,m}, \mathbf{C}_{l,m}\}$:

$$\begin{aligned} y(s_i) = \sum_{l=1}^{L}\sum_{m=-l}^{l}\Big[&\{\Re(f^B_{l,m})\Re(\mathbf{B}_{l,m}(s_i)) - \Im(f^B_{l,m})\Im(\mathbf{B}_{l,m}(s_i))\} \\ &+ \{\Re(f^C_{l,m})\Re(\mathbf{C}_{l,m}(s_i)) - \Im(f^C_{l,m})\Im(\mathbf{C}_{l,m}(s_i))\}\Big] + \epsilon_i. \end{aligned} \tag{2.21}$$

The model equation in matrix form and its covariance structure are discussed in Subsection 2.3.2.

2.3.2. TVRF in Matrix Form. We denote the n observations on the TVRF by the $2n \times 1$ vector $\mathbf{y} = [y^\theta(s_i)_{1\leq i\leq n},\ y^\phi(s_i)_{1\leq i\leq n}]^T$, where y^θ is the components of θ (co-latitude) direction and y^ϕ is the component of the ϕ (longitude) direction.

The TVRF model in the matrix form is thus expressed as

$$\mathbf{y} = \mathbf{Z}\mathbf{f} + \boldsymbol{\epsilon}, \tag{2.22}$$

where $\mathbf{Z}$ is a $2n \times p$ matrix, including with evaluations of VSH at locations, $s_i \in \mathbb{S}^2$. $\mathbf{f}$ is a length p vector of the random VSH coefficients. The number of the coefficients, p, depends on the maximum

degree of VSH so that $p = 2(L^2 + 2L) + 2(L^2 + L)$. The coefficient vector, $\mathbf{f}$, is defined as follows:

(2.23)
$$\mathbf{f} = \big(\underbrace{\Re(f^B_{1,.}), \Re(f^B_{2,.}), \cdots \Re(f^B_{L,.})}_{\text{Real part of } f^B_{l,m}}, \underbrace{\Im(f^B_{1,.}), \cdots \Im(f^B_{L,.})}_{\text{Imaginary of } f^B_{l,m}}, \underbrace{\Re(f^C_{1,.}), \cdots \Re(f^C_{L,.})}_{\text{Real of } f^C_{l,m}}, \underbrace{\Im(f^C_{1,.}), \cdots, \Im(f^C_{L,.})}_{\text{Imaginary of } f^C_{l,m}}\big)^T$$

where $\Re(f^\dagger_{l,.})$ is the vector of real parts of $[f^\dagger_{l,m}]_{-l\le m\le l}$, and $\Im(f^\dagger_{l,.})$ is the vector of imaginary parts of $[f^\dagger_{l,m}]_{-l\le m<0,0<m\le l}$ for a fixed l. Finally, $\boldsymbol{\epsilon}$ is a $2n \times 1$ vector of noise terms with i.i.d. $N(0, \tau^2)$ coordinates.

$\mathbf{Z}$ consists of evaluations of VSH at locations $\{s_i\}_{i=1}^n$. The structure of $\mathbf{Z}$ is described as follows.

(2.24)
$$\mathbf{Z} = \begin{pmatrix} \Re(\mathbf{B}^\theta_{l,m}) & -\Im(\mathbf{B}^\theta_{l,m}) & \Re(\mathbf{C}^\theta_{l,m}) & -\Im(\mathbf{C}^\theta_{l,m}) \\ \Re(\mathbf{B}^\phi_{l,m}) & -\Im(\mathbf{B}^\phi_{l,m}) & \Re(\mathbf{C}^\phi_{l,m}) & -\Im(\mathbf{C}^\phi_{l,m}) \end{pmatrix}.$$

$\mathbf{B}^\theta_{l,m}, \mathbf{C}^\theta_{l,m}$ are the θ components on the bases, and $\mathbf{B}^\phi_{l,m}, \mathbf{C}^\phi_{l,m}$ are the ϕ components. The evaluations of VSH for each l, m, are arranged in $\mathbf{Z}$ in accordance with the l, m order of $\mathbf{f}$ as defined in (2.23).

The vector $\mathbf{y}$ is normally distributed with a mean of $\mathbf{0}$ and variance $V = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \tau^2 I_{2n}$ where $\mathbf{G}$ is a $p \times p$ covariance matrix of the random coefficient vector, $\mathbf{f}$, and I_{2n} is the $2n \times 2n$ identity matrix. G is a block matrix, and its diagonal blocks are the covariance matrices of the subvector $f^\dagger_l$ of $\mathbf{f}$, for $\dagger \in \{B, C\}$ and $l = 1, 2, \ldots, L$.

(2.25)
$$\mathbf{G} = \text{diag}\Big[\big[Var\{\Re(f^B_{l,.})\}\big]_{1\le l\le L}, \big[Var\{\Im(f^B_{l,.})\}\big]_{1\le l\le L}, \big[Var\{\Re(f^C_{l,.})\}\big]_{1\le l\le L}, \big[Var\{\Im(f^C_{l,.})\}\big]_{1\le l\le L}\Big]$$

with each covariance matrix, $Var\{\Re(f^\dagger_{l,.})\}, Var\{\Im(f^\dagger_{l,.})\}$ defined as

$$Var\{\Re(f^\dagger_{l,.})\} = \frac{1}{2}\sigma^2_{\dagger,l}(I_{2l+1} + J_{2l+1}),$$
$$Var\{\Im(f^\dagger_{l,.})\} = \frac{1}{2}\sigma^2_{\dagger,l}(I_{2l} + J_{2l}),$$

and J_{2l+1} and J_{2l} are defined as

$$J_{2l+1} = \text{anti-diag}[(-1)^{|l|}, (-1)^{|l-1|}, \cdots, (-1)^0, \cdots, (-1)^{|l-1|}, (-1)^{|l|}],$$
$$J_{2l} = \text{anti-diag}[(-1)^{|l+1|}, (-1)^{|l|}, \cdots, (-1)^{|2|}, (-1)^{|2|}, \cdots, (-1)^{|l|}, (-1)^{|l+1|}].$$

Here, the anti-diagonal matrix anti-diag$[a_1, a_2] = \begin{pmatrix} 0 & a_1 \\ a_2 & 0 \end{pmatrix}$. The J matrices ensure the conjugate relationship that $Cov\{\Re(f^{\dagger}_{l,-m}), \Re(f^{\dagger}_{l,m})\} = (-1)^{|m|} Var\{\Re(f^{\dagger}_{l,m})\}$, $Cov\{\Im(f^{\dagger}_{l,-m}), \Im(f^{\dagger}_{l,m})\} = (-1)^{|m+1|} Var\{\Im(f^{\dagger}_{l,m})\}$.

We also model the variance parameters $\sigma^2_{\dagger,l}$, $\dagger \in \{B, C\}$ as follows:

$$\sigma^2_{\dagger,l} = \sigma^2_{\dagger} l^{-\alpha_{\dagger}}, \ \dagger \in \{B, C\}, \tag{2.26}$$

where $\alpha_B, \alpha_C \geq 0$, and $\sigma^2_B, \sigma^2_C > 0$. Notice that if $L = \infty$, then in order for the process X to have finite variance, we must have $\alpha_B > 1$ and $\alpha_C > 1$. The parameters (α_B, α_C) control the smoothness of the process since a relatively large value of these parameters implies that only the VSH functions with corresponding to small frequency index l plays a dominant role in the description of the process.

2.3.3. Dependency: Cross-correlation between Random Coefficients. We introduce dependencies between random coefficients of curl-free fields and divergence-free components of the TVRF $X(\cdot)$ by introducing correlation among the coefficients in the VSH representation. For a fixed l, we impose a correlation, $\rho_l = corr(f^B_{l,m}, f^C_{l,m})$ for $|m| \leq l$. In the dependent case, let $\mathbf{G}_r$ be the covariance of random coefficients. Then $V_r = \mathbf{Z}\mathbf{G}_r\mathbf{Z}^T + \tau^2 I_{2n}$ is the covariance matrix of the observation vector $\mathbf{y}$. The covariance matrix, $\mathbf{G}_r$, is defined as

$$\mathbf{G}_r = \begin{pmatrix} G_B & G_{B,C} \\ G_{C,B} & G_C \end{pmatrix},$$

where

$$\begin{aligned} G_{\dagger} &= \text{diag}\Big[\big[Var\{\Re(f^{\dagger}_{l,.})\}\big]_{1\leq l\leq L}, \big[Var\{\Im(f^{\dagger}_{l,.})\}\big]_{1\leq l\leq L}\Big], \\ G_{B,C} &= \text{diag}\Big[\big[Cov\{\Re(f^{B}_{l,.}), \Re(f^{C}_{l,.})\}\big]_{1\leq l\leq L}, \big[Cov\{\Im(f^{B}_{l,.}), \Im(f^{C}_{l,.})\}\big]_{1\leq l\leq L}\Big], \\ &= \{G_{C,B}\}^T \end{aligned}$$

with

$$Cov\{\Re(f^B_{l,\cdot}), \Re(f^C_{l,\cdot})\} = \frac{1}{2}\rho_l \sigma_{B,l}\sigma_{C,l}(I_{2l+1} + J_{2l+1}),$$
$$Cov\{\Im(f^B_{l,\cdot}), \Im(f^C_{l,\cdot})\} = \frac{1}{2}\rho_l \sigma_{B,l}\sigma_{C,l}(I_{2l} + J_{2l}).$$

We additionally assume that ρ_l decreases as l increases. A particular choice is $\rho_l = \rho l^{-\beta}$, for $1 \le l \le L_\rho$, and $\rho_l = 0$ for $l > L_\rho$, for some $\rho \in (-1, 1)$, $\beta \ge 0$ and $L_\rho \le L$. β can be seen as a smoothness parameter for the correlation.

2.3.4. Extension of the TVRF Model. We extend the stationary TVRF modeling to model spatio-temporal data that arise in the context of geomagnetism, by relaxing the requirement that the TVRF is componentwise isotropic. We also allow time dependencies among the random coefficients, by assuming that the mean fo the (l, m)-th coefficient at time t, namely, $\mathbb{E}(f^\dagger_{l,m}(t))$ equals $\sum_{j=1}^J b^\dagger_{l,m,j}(t)\mathcal{B}_j(t)$, where $\{\mathcal{B}_j(t)\}_{j=1}^J$ is a collection of smooth, linearly independent function, defined on the time domain. In the subsequent applications, we take $\{\mathcal{B}_j(t)\}_{j=1}^J$ to be a B-spline basis. We define a scaling function, $\varphi(s)$, to cope with an anisotropic process, multiplied to the isotropic TRVF model, (2.19) (2.22), to transform into an anisotropic field, such that

$$\mathbf{y} = \boldsymbol{\varphi} \otimes \big(\mathbf{Zf} + \boldsymbol{\epsilon}\big), \tag{2.27}$$

where $\boldsymbol{\varphi} = (\varphi^{\boldsymbol{\theta}}(s_i), \varphi^{\boldsymbol{\phi}}(s_i))_{i=1}^n$ is a $2n \times 1$ vector, and $\otimes$ denotes coordinatewise multiplication. The covariance matrix can be updated to $V = \Phi(\mathbf{ZGZ}^T + \tau^2 I)\Phi$ or $V_r = \Phi(\mathbf{ZG}_r\mathbf{Z}^T + \tau^2 I)\Phi$, where Φ is the $2n \times 2n$ diagonal matrix with the vector of diagonal elements being $\boldsymbol{\varphi}$. The scaling functions for the co-latitude (θ) or longitude (ϕ) directions can either be the same, i.e., $\varphi^{\boldsymbol{\theta}}(s_i) = \varphi^{\boldsymbol{\phi}}(s_i)$, or different, $\varphi^{\boldsymbol{\theta}}(s_i) \ne \varphi^{\boldsymbol{\phi}}(s_i)$, depending on data features. We discuss the details and illustrate its application to modeling Earth's magnetic field in Section 2.6.

2.4. Statistical Estimation and Prediction

We describe the computation of the maximum likelihood estimate (MLE) of the model parameters and predictions for the proposed TVRF model. We discuss conditions for the consistency of the MLEs and present confidence sets of the parameters based on the asymptotic behavior of the

MLE, treating L, the maximum order of VSH (henceforth, referred to as *model order*), as known. We discuss the selection of the model order L and the process smoothness parameters α_B, α_C, and β by AIC, AICc, or BIC criteria.

2.4.1. Log-Likelihood Functions. We start with the construction of the log-likelihood function of TVRF on a sphere. We assume that we observe a spatio-temporal data set in which there are n_t observations at time t, where $t = 1, \cdots, T$. At locations $s_{t,i} = (\theta_{t,i}, \phi_{t,i}) \in \mathbb{S}^2,\ i = 1, \cdots, n_t$, we observe the noisy tangential vector field at time t as $\mathbf{y}_t = \big[y_t(s_{t,1}), y_t(s_{t,2}), \cdots, y_t(s_{t,n_t})\big]^T$. $\mathbf{y}_t$ follows a Gaussian distribution with a mean of $\mathbf{0}$ and the covariance, $V_t(\boldsymbol{\theta}) = \mathbf{Z}_t\mathbf{G}\mathbf{Z}_t^T + \tau^2 I_{2n_t}$, where $\boldsymbol{\theta} = [\sigma_B^2, \sigma_C^2, \rho, \tau^2]$, by treating the model order L, and the decay parameters $(\alpha_B, \alpha_C, \beta)$ as known, or given. With an assumption of independence and identical distribution across time, the log-likelihood of the observed data is

$$\ell(\boldsymbol{\theta}) = -\frac{\sum_{t=1}^T n_t}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\Big[\log(|V_t(\boldsymbol{\theta})|) + (\mathbf{y}_t^T V_t(\boldsymbol{\theta})^{-1}\mathbf{y}_t)\Big]. \tag{2.28}$$

The log-likelihood is simplified if the observations are located at identical positions across time t. Then the number of observations per time period is the same, i.e, $n_t \equiv n$, for all t. $\mathbf{Z}_t$, the matrix of evaluations of VSH, at the locations are identical across time, so we set $\mathbf{Z} = \mathbf{Z}_t$ for all t. The covariance matrix becomes $V(\boldsymbol{\theta}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \tau^2 I_{2n}$. The log-likelihood of the observations of the identical locations is

$$\ell(\boldsymbol{\theta}) = -\frac{nT}{2}\log(2\pi) - \frac{1}{2}\Big[T\cdot\log\big(|V(\boldsymbol{\theta})|\big) + \mathrm{trace}\big(V(\boldsymbol{\theta})^{-1}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t^T\big)\Big]. \tag{2.29}$$

Score Equations. The first and second derivatives of $\ell(\boldsymbol{\theta})$ (2.28), (2.29) with respect to each parameter, are, for each time point t (suppressing the dependence on t), given by

$$\begin{aligned}
\frac{\partial\ell(\boldsymbol{\theta})}{\partial\theta_i} &= -\frac{1}{2}\mathrm{tr}(V^{-1}\frac{\partial V}{\partial\theta_i}) + \frac{1}{2}\mathbf{y}^T V^{-1}\frac{\partial V}{\partial\theta_i}V^{-1}\mathbf{y},\\
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} &= \frac{1}{2}\mathrm{tr}(V^{-1}\frac{\partial V}{\partial\theta_j}V^{-1}\frac{\partial V}{\partial\theta_i}) - \frac{1}{2}\mathbf{y}^T V^{-1}\frac{\partial V}{\partial\theta_i}V^{-1}\frac{\partial V}{\partial\theta_j}V^{-1}\mathbf{y} - \frac{1}{2}\mathbf{y}^T V^{-1}\frac{\partial V}{\partial\theta_j}V^{-1}\frac{\partial V}{\partial\theta_i}V^{-1}\mathbf{y},
\end{aligned}$$

where $\theta_i \in \boldsymbol{\theta},\ i, j = 1, 2, 3, 4$, and $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, \theta_4] = [\sigma_B^2, \sigma_C^2, \rho, \tau^2]$.

The score functions and their derivatives are complicated functions involving inverses of large dimensional matrices, and the solution of the score equation is not available in closed form. For this

reason, we maximize the log-likelihood function through numerical optimization in order to find MLEs. We conduct the optimization using the Nelder-Mead method for $\sigma_B^2, \sigma_C^2, \tau^2$ and the BFGS quasi-Newton method with a box constraint for ρ, which is implemented by using the `optim()` function in `R`.

2.4.2. Initial Values for Optimization of the Likelihood. Good initial values for optimization enhance the accuracy of estimates and reduce their computational cost. Utilizing our model assumptions, we develop a method for finding good initial values for parameters for maximizing the log-likelihood function.

Variances of Random Coefficients. The framework of obtaining initial estimators of the variances of the random coefficients is derived from the calculation of sample variances. This method requires the VSH coefficients to be known beforehand. The VSH coefficients are calculated using the normal equation $\hat{\mathbf{f}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$ or discrete VSH transforms if the observations are on an equiangular grid.

We denote the initial estimator as $\sigma^2_{\dagger,ini}$ for $\sigma^2_\dagger$, for a given smoothness parameters, $\alpha_\dagger$ (2.20). We start to compute the initial estimator of lth frequency variance, $\sigma^2_{\dagger,l,ini}, l = 1, 2, \cdots, L$, as follows.

$$\sigma^2_{\dagger,l,ini} = \frac{1}{2l} \sum_{0\le|m|\le l} \left(\Re(\widehat{f}^\dagger_{l,m}) - \Re(\overline{f}^\dagger_{l,\cdot})\right)^2 + \frac{1}{2l-1} \sum_{1\le|m|\le l} \left(\Im(\widehat{f}^\dagger_{l,m}) - \Im(\overline{f}^\dagger_{l,\cdot})\right)^2, \tag{2.30}$$

where $\overline{f}^\dagger_{l,\cdot}$ is the arithmetic mean of $\widehat{f}^\dagger_{l,m}$. Given the smoothness parameter, we express $\sigma^2_{\dagger,\text{ini}}$ the form of $\sigma^2_{\dagger,l,\text{ini}}$.

$$L\sigma^2_{\dagger,ini} = \sum_{l=1}^{L} l^{\alpha_\dagger}\sigma^2_{\dagger,l,ini}. \tag{2.31}$$

If $\alpha_\dagger$ is known, we can use Equation (2.31) to obtain the initial estimators, $\sigma^2_{\dagger,ini}$.

$$\sigma^2_{\dagger,ini} = \frac{1}{L}\sum_{l=1}^{L} \sigma^2_{\dagger,l,ini} l^{\alpha_\dagger}. \tag{2.32}$$

Alternatively, we may use linear regression, after noticing the following linear relationship after a logarithmic transformation of $\sigma^2_{\dagger,l}$.

$$\log \sigma^2_{\dagger,l,ini} = \log \sigma^2_{\dagger,ini} - \alpha_\dagger \log l, \text{ where } l = 2, 3, \cdots, L \tag{2.33}$$

The regression method is implemented by using the regression model (2.33) with $\log l$ as the explanatory variable and $\log \sigma^2_{\dagger,l,ini}$ as the response variable, and we exclude first few l, (here $l = 1$). The exponential of the estimated intercept is chosen to be our initial estimate, $\sigma^2_{\dagger,ini}$ and the slope suggests a crude estimate of the smoothness parameter, $\alpha_\dagger$ for $\dagger \in \{B, C\}$.

Correlation between Random Coefficients. We define the correlation $\rho_l = corr(f^B_{l,m}, f^C_{l,m})$, $l = 1, 2, \cdots, L_\rho$ with the decay parameter β as $\rho_l = \rho l^{-\beta}$. We express ρ as a weighted sum of ρ_l.

$$L_{\rho_{ini}} = \sum_{l=1}^{L_\rho} l^\beta \rho_{l,ini} \tag{2.34}$$

We calculate $\rho_{l,ini}$ from empirical correlations from estimated $f^B_{l,m}, f^C_{l,m}$. If β is known, we can use (2.34) directly to obtain the initial estimate ρ_{ini}. Alternatively, we may use the linear regression of $\log |\rho_{l,ini}|$ on $\log l$, (in analogy with the regression model (2.33)) to obtain the initial estimate of $|\rho|$, namely, $|\rho_{ini}|$ as the exponential of the intercept, and a crude estimate of β as the negative of the slope. The sign of ρ_{ini} is chosen as $sign(\sum_{l=1}^{L_\rho} l^{\hat{\beta}} \rho_{l,ini})$.

Variances of Random Errors. We use the variance of residuals as an initial estimator for τ^2, where the residuals are defined as $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$. Accordingly, the initial estimator, τ^2_{ini}, is $Var(\mathbf{e})$. If observations are on an equiangular grid, we use the structure of the grid. Let $y(s_c)$ be an observation located at a "center location" s_c, and $y(s_{n(c)})$ be the observation at a randomly chosen location $s_{n(c)}$ among the neighboring locations of the center point. The initial estimator is derived from the relationship that $Var\{y_(s_c) - y(s_{n(c)})\} \geq 2\tau^2$. Using this lower bound, and assuming a reasonably dense sampling scheme, we contend that half of the estimated variance $Var\{y_(s_c) - y(s_{n(c)})\}$ would be a meaningful initial estimator for τ^2. Given a 20×20 equiangular grid, we have 36 of pairs, $(y(s_c), y(s_{n(c)}))$, excluding selections from one of the two boundaries, $(\theta_i, \phi_{20}), (\theta_{20}, \phi_i), i = 1, \cdots, 20$, in order to prevent duplicate selections.

As an alternative, the estimation method of noise variance in the context of two-dimensional signal processing by Hall et al. [16] can also be used to obtain an initial estimate on an equiangular grid. We apply this method by selecting 2×2 squares in the grid and use the mean of the squared weight sum of observations over each square as an initial estimate. The weights on the nodes of the squares are $(-3a, a, a, a)$, where $a = 12^{-1/2}$. The resulting $\hat{\tau}^2_{ini}$ can be seen as the variance of the residuals based on an implicit prediction scheme of the observation y at the corner of each square, based on the data at the neighboring nodes of the squares.

2.4.3. Selections of L and Smoothness Parameters. The optimization of likelihood cannot be used effectively to determine the model order L and L_ρ. Maximization of likelihood also does not provide very accurate estimates of the smoothness parameters, $\{\alpha_B, \alpha_C, \beta\}$, unless the sample size is quite large. Thus, we treat the selection of L as a model selection problem. Furthermore, due to limited sensitivity of the likelihood to the smoothness parameters, we estimate them separately, while keeping their values fixed in the maximization of the likelihood with respect to $\boldsymbol{\theta} = (\sigma^2_B, \sigma^2_C, \rho, \tau^2)$. We evaluate the Akaike information criterion (AIC), corrected AIC (AICc), and Bayesian Information criterion (BIC), and compare the predicted mean squared error (PMSE, out-of-sample errors) to select L after obtaining the MLEs of $\boldsymbol{\theta}$ for different values of the model order L. All methods recommend a reasonable L in our isotropic simulations, and BIC and PMSE recommend a reasonable L in anisotropic simulations in Section 2.5.

As we indicated earlier, the smoothness parameters control the contributions of the VSH basis functions corresponding to different harmonic frequency l to the overall fluctuation of the TVRF. A larger α indicates that the first few VSH are dominant, and the process is very smooth. The smoothness with the deterministic model helps predictions. However, due to limited sensitivity of the process to the changes in the smoothness parameter value, over a fairly wide range, the smoothness parameters are inherently difficult to estimate based on a limited amount of data. We compute PMSE against different smoothness parameters and compare the results with one of the true parameters in the simulated data set. The profiles of PMSE against α, including the true parameters, are flat, which reinforces the point that identifying α is a challenge. A smaller α implies many fluctuations across more l, and it produces difficulties in predicting the model. We

either use crude estimates of the smoothness parameters using the regression method (2.33) or by eye-balling of the plots the coefficients $\sigma^2_{\dagger,l,ini}$ against l to judge the decay ratio.

2.4.4. Asymptotic Property of MLEs. Here we briefly indicate the conditions under which the MLE of the parameter $\boldsymbol{\theta}$ is consistent as the model order, L, increases to infinity. We denote the true parameters as $\boldsymbol{\theta_0} = (\sigma_B^2, \sigma_C^2, \rho, \tau^2)$. We assume the smoothness parameters α_B, α_C and β are fixed.

When $T = 1, N = 2n$, the log-likelihood equation (2.28) becomes

$$\ell(\boldsymbol{\theta}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\big[\log(|\mathbf{ZG}(\boldsymbol{\theta})\mathbf{Z}^T + \tau^2\mathbf{I}_N|) + \boldsymbol{y}^T(\mathbf{ZG}(\boldsymbol{\theta})\mathbf{Z}^T + \tau^2\mathbf{I}_N)^{-1}\boldsymbol{y}\big]. \tag{2.35}$$

Using the theory of linear random effects modeling, we can show that, under appropriate regularity conditions, the MLE is consistent and asymptotically normal with the asymptotic covariance which is the inverse of the Fisher information. The key regularity condition needed to prove this result is the increasing denseness of the sampling design, as $L \to \infty$ slowly with n. For i.i.d. realizations of the process observed across times $t = 1, \ldots, T$, in order to prove consistency, it suffices to have $T \to \infty$ together with $L \to \infty$, as long as the observation locations across times are independently and randomly distributed on $\mathbb{S}^2$. If the data are on a regular grid for all time points, then we still need the grid size to increase to infinity with L, while T may even be bounded.

2.4.5. Predictions of the TVRF. We predict the tangential vector field by making use of the theory of linear mixed models. Specifically, we predict the TVRF at any new location by its Best Linear Unbiased Predictor (BLUP). The BLUPs, $\widetilde{\mathbf{y}}_{new}$, of new locations given the observed locations are defined as

$$\widetilde{\mathbf{y}}_{new} = \mathbb{E}(\mathbf{Z}_{new}\mathbf{f}|\mathbf{y}_{old}) = \mathbf{Z}_{new}\widehat{\mathbf{G}}\mathbf{Z}_{old}^T\mathbf{V}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{y}_{old}, \tag{2.36}$$

where $\mathbf{V}(\hat{\boldsymbol{\theta}}) = \mathbf{Z}_{old}\widehat{\boldsymbol{G}}\mathbf{Z}_{old}^T + \hat{\tau}^2 I$. $\mathbf{Z}_{old}$ are the VSH evaluations of the observed locations, and $\mathbf{Z}_{new}$ are the evaluations of the new locations.

2.4.6. Confidence Set. We construct confidence sets by combining the asymptotic theory and parametric bootstrap sampling. The asymptotic theory suggests a confidence set for the parameter

$\boldsymbol{\theta}$ by making use of asymptotic distribution of the following Wilks-type statistics,

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T (\sigma^2(\hat{\boldsymbol{\theta}}))^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \frac{p(n-1)}{n-p} F_{p,n-p},$$

where $\hat{\boldsymbol{\theta}}$ is the MLE for $\boldsymbol{\theta}$ and $\sigma^2(\hat{\boldsymbol{\theta}})$ is the asymptotic variance of $\hat{\boldsymbol{\theta}}$. Since $\sigma^2(\hat{\boldsymbol{\theta}})$ is not available in closed form, and even its approximation involves high-dimensional matrix inversions, as an alternative, we adopt a parametric bootstrapping approach. We simulate bootstrap samples from estimated distributions using the MLE $\hat{\boldsymbol{\theta}}$ as the true parameter. Let $\hat{\boldsymbol{\theta}}^*$ be the MLE from the bootstrapped data. Using the asymptotic theory of Wilks-type statistics, we may construct the bootstrap confidence set for $\boldsymbol{\theta}$ as

$$(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})^T \left(\widehat{\sigma^2(\hat{\boldsymbol{\theta}})} \right)^{-1} (\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}) \leq \frac{p(n-1)}{n-p} F_{p,n-p}(1-\alpha).$$

Then we define the $(1-\alpha)100\%$ level confidence set for the different coordinates of $\boldsymbol{\theta}$ as

$$\theta_i \in \left[\hat{\theta}_i - \sqrt{q^*_{1-\alpha}} \cdot \widehat{\sigma(\hat{\boldsymbol{\theta}})}_i, \hat{\theta}_i + \sqrt{q^*_{1-\alpha}} \cdot \widehat{\sigma(\hat{\boldsymbol{\theta}})}_i \right], \tag{2.37}$$

where $q^*_{1-\alpha} = \frac{p(n-1)}{n-p} F_{p,n-p}(1-\alpha)$. $\widehat{\sigma^2(\hat{\boldsymbol{\theta}})}$ is the sample covariance matrix of the bootstrap estimates and $\widehat{\sigma^2(\hat{\boldsymbol{\theta}})}_i$ is ith diagonal element of the sample covariance matrix. We consider the *bias-corrected and acceleration bootstrap* (BC_a) confidence interval [9]. If we create a confidence interval for a single parameter θ_i, the $(1-\alpha)100\%$ confidence interval is $\theta_i \in \left[\hat{\theta}_i - z^{(\alpha/2)} \hat{\sigma}_i, \hat{\theta}_i + z^{(1-\alpha/2)} \hat{\sigma}_i \right]$, where $\hat{\sigma}$ is the standard deviation of the estimator $\hat{\theta}_i$. $z^{(\alpha)}$ is the $100 \cdot \alpha$ percentile point of a standard normal variate. Under some regularly conditions, we set the correct $(1-\alpha)100\%$ confidence interval for θ as

$$\theta_i \in \left[\hat{F}^{-1}(\Phi(z[\alpha/2])), \hat{F}^{-1}(\Phi(z[1-\alpha/2])) \right], \tag{2.38}$$

where Φ denotes the standard normal cdf and $\hat{F}^{-1}(\alpha)$ is the $100 \cdot \alpha$ percentile of bootstrap CDF, $\hat{F}(s) = Pr_\theta\{\hat{\theta}^* < s\}$. $z[\alpha] = z_0 + \frac{z_0 + z^{(\alpha)}}{1 - a(z_0 + z^{(\alpha)})}$ The bias-correction constant z_0 is $z_0 = \Phi^{-1}(\hat{F}(\hat{\theta}))$. The acceleration constant a is $a = \frac{1}{6}\left[\frac{\mu_3(X)}{\mu_2(X)^{3/2}}\right] = \frac{1}{6}\left[\frac{E[(X-\mu)^3]}{E[(X-\mu)^2]^{3/2}}\right]$.

The other approach to building confidence intervals is using the percentiles directly from the bootstrap replicates. Given B bootstrap replicates of $\hat{\theta}$, we define the $(1-\alpha)$ level of the confidence

interval as

$$\theta_i \in \left[\hat{\theta}^*_{i(\alpha/2\cdot B)}, \hat{\theta}^*_{i((1-\alpha/2)\cdot B)}\right], \tag{2.39}$$

where $\hat{\theta}^*_{i(j)}$ is jth position when we arrange the bootstrap estimates of θ_i in ascending order.

We demonstrate the construction of confidence intervals in Section 2.5.

2.5. Simulation Studies

In this section, we conduct numerical studies to illustrate the estimation, model selection and prediction performances of the TVRF. We explore how sample size and L are connected to the estimation performance and examine model selection methods. We consider the extended TVRF model with a non-zero and possibly time-varying mean field and axially symmetric variance profiles for the different components of the random field.

2.5.1. Setting. We first consider the Gaussian isotropic TVRF models with $f^B_{l,m} \sim CN(0, \sigma^2_B)$, $f^C_{l,m} \sim CN(0, \sigma^2_C)$, and $\epsilon \sim N(0, \tau^2)$. We set its parameters $\sigma^2_B = 20, \sigma^2_C = 10, \tau^2 = 0.01$ with fixed smoothness parameters $\alpha_B = 2, \alpha_C = 2$. The observation, $y(s_i)$ at location $s_i \in \mathbb{S}^2$ is as follows.

$$y(s_i) = \sum_{l=1}^{L} \sum_{m=-l}^{l} \left\{ f^B_{l,m} \mathbf{B}_{l,m}(s_i) + f^C_{l,m} \mathbf{C}_{l,m}(s_i) \right\} + \epsilon_i, \text{ where } i = 1, 2, \cdots, n. \tag{2.40}$$

The locations $s_i \in \mathbb{S}^2$ are randomly selected or on an equiangular grid. The locations $s_i \equiv (\theta_j, \phi_k)$ are on the $2B \times 2B$ equiangular grid defined as $\theta_j = \pi(2j+1)/4B, \phi_k = 2\pi k/2B$, where $j, k = 0, 1, \cdots, 2B-1$ with a given bandwidth B. *Nyquist-Shannon sampling theorem* for a sphere [29] suggests that the maximum L is equal to B in the discrete SSH transforms. The VSH transform requires one higher harmonic frequency for the associated SSH transforms, so the maximum harmonic frequency for the associated L for VSH is $B-1$ on the $2B \times 2B$ grid resulting in $n = (2B)^2$ locations indexed by the pairs (θ_j, ϕ_k). We use $L = 9$ and the sample size is 400 for the equiangular grid. For comparison, we also consider a random design with the same number of points. For random locations, we choose n pairs, (v_i, u_i) from a uniform distribution with minimum 0 and maximum 1 separately and converted the pair to locations, (θ_i, ϕ_i), in spherical coordinates where

$\theta_i = \arccos(2v_i - 1), \phi_i = 2\pi u_i$. The configuration of points plays an important role in our model fitting because it is based on evaluations of VSH at the locations.

We use 500 replicates of the isotropic TVRFs on identical locations for performance evaluation.

2.5.2. Independent Cases.

Sample Size and L. We investigate what sample sizes are required for the maximum degree of VSH based on the sampling theorem to estimate the parameter properly. We consider the cases at $L = 9$ or $L = 19$ with equiangular locations or random locations, with the corresponding sample sizes being 400 or 1600. We estimate parameters with known L and α_B, α_C. The summary statistics of the estimation are given in Table 2.1. For the equiangular grid, the spread of estimates is centered at the true values of the parameters, and a larger sample size produces a smaller dispersion. For the case of randomly distributed locations, the performance of the estimates is more dependent on sample size. The results of the estimation over the data set generated by $L = 19$ and $n = 400$ random locations produce biased and highly variable outcomes, though the results over the data set generated by $L = 19$ and $n = 1600$ random locations result in estimates close to the true parameters. This simulation study suggests that the maximum model order L should be specified based on the available sample size n. As n increases, we can have more precise estimates when L is known. There is no significant difference between the equiangular grid and random points if the sample size satisfies the key condition of the Sampling Theorem: $(n \geq (2(L+1))^2)$.

Estimation and Prediction. We simulate data from the model (2.40). We allow extensions of the model to have nonzero mean fields and location-dependent variability of the random field modeled by (2.27). We refer to the latter model as the *anisotropic TVRF* model, even though the form of the anisotropy in (2.27) is rather specific, since the mean-centered filed is an isotropic TVRF except for a pointwise multiplication by an axially symmetric non-negative function (standard deviation profile, or scaling function) φ. We generate three cases: isotropic TVRF with a mean of zero, isotropic TVRF with a mean of μ, and anisotropic TVRF with a mean of μ. The three TVRFs

True $\sigma_B^2 = 20$

Case	L	n	Min.	Q1	Median	Mean	Q3	Max.
Equiangular	9	400	12.639	17.533	19.373	19.268	20.947	27.332
Random	9	400	12.733	17.498	19.430	19.266	20.933	27.309
Equiangular	19	400	16.172	18.758	19.675	19.884	20.946	24.073
Random	19	400	29.839	75.354	89.086	90.101	103.781	134.116
Equiangular	19	1600	17.530	19.032	19.844	20.009	20.826	23.758
Random	19	1600	17.492	19.049	19.845	20.004	20.816	23.773

True $\sigma_C^2 = 10$

Case	L	n	Min.	Q1	Median	Mean	Q3	Max.
Equiangular	9	400	6.302	9.141	10.349	10.134	11.141	14.431
Random	9	400	6.260	9.098	10.330	10.125	11.122	14.460
Equiangular	19	400	7.675	9.217	9.817	9.883	10.391	12.537
Random	19	400	21.140	65.858	79.463	81.012	93.014	125.712
Equiangular	19	1600	7.785	9.512	9.912	9.951	10.511	11.276
Random	19	1600	7.704	9.543	9.902	9.952	10.494	11.324

TABLE 2.1. Summary statistics of estimates $(\hat{\sigma}_B^2, \hat{\sigma}_C^2)$: L is the maximum order of VSH. n is the sample size.

that we consider for this simulation study are expressed as follows.

$$Y_1(s) = y(s) + \epsilon, \tag{2.41}$$

$$Y_2(s) = \mu(s) + y(s) + \epsilon, \tag{2.42}$$

$$Y_3(s) = \mu(s) + \varphi(\theta(s))\big(y(s) + \epsilon\big), \tag{2.43}$$

where $\mu(s)$ is the mean at location $s \in \mathbb{S}^2$, $y(s)$ is TVRF observation at s, and $\theta(s)$ denotes the θ coordinate of the location s. Finally, ϵ's are i.i.d Gaussian noise with zero mean and coordinatewise variance τ^2. We assume a fixed mean field represented in the VSH basis such that $\mu(s) = \sum_{l=1}^{L_\mu}\{h_{l,m}^B \mathbf{B}_{l,m}(s) + h_{l,m}^C \mathbf{C}_{l,m}(s)\}$ where the maximum degree of VSH for the mean field is 3 ($L_\mu = 3$). The fixed coefficients, $\{h_{l,m}^B, h_{l,m}^C\}$, are complex numbers and satisfy the conjugacy relationship. TVRFs are generated with an identical setting from Equation (2.40), where the maximum degree of VSH is 9 ($L_R = 9$). We allow a higher value of the model order L_R (the subscript R refers to the "residual field", i.e., after subtracting off the mean) for the isotropic TVRF y than the maximum degree L_μ for the mean field to reflect that residual fields have more fluctuations than the mean field.

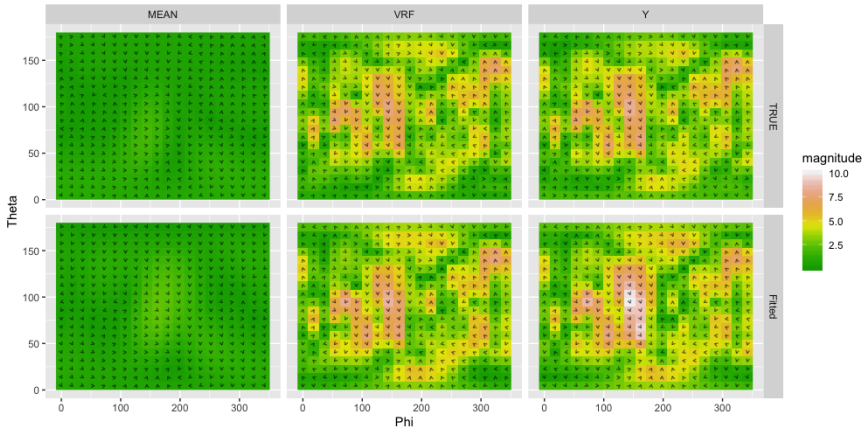


FIGURE 2.3. Isotropic field with mean field: Top presents true fields, and bottom presents fitted fields. VRF indicates isotropic TVRF. Y are observed fields $Y_2(s)$

We link the mean field μ to the standard deviation profile φ in the description of the anisotropic field given by (2.43) since this description is closely connected to our application on the Ørsted geomagnetic data set. Accordingly, we model a scaling function, $\varphi(\theta)$, as $\sqrt{\int_\phi (\mu(\theta,\phi)^2) d\phi}$ separately for each of the two canonical coordinate directions of the tangential vector field (θ = co-latitude, and ϕ = longitude). Each $\varphi(\theta)$ is dependent on co-latitude (θ) only, so the anisotropic fields described by (2.43) are axially symmetric.

We generate 500 replicates of the three models with 15 independent time periods (500 replicates of 15 independent realizations of the processes at the same locations). In the $\mu \equiv 0$ case, we estimate TVRF parameters using maximum likelihood, based on the combined data set consisting of 15 temporally independent realizations of the processes, treating smoothness parameters ($\alpha_B = \alpha_C = 2$) and the model order for the TVRF L_R as known, and evaluate the prediction performance. In the cases of nonzero mean fields, $\mu(s)$, we fit the mean field first by least squares regression, ignoring the correlation structures, and thereafter fit the parameters for the isotropic/anisotropic TVRF model by using the data from the estimated mean-subtracted residual field. For estimation of the mean field, we regress the observed data against the evaluation of VSH with given maximum order $L_\mu = 3$, and least squares estimates of the fixed coefficients, $\{h^B_{l,m}, h^C_{l,m}\}$. In the isotropic case, we conduct TVRF modeling onto the residual fields, $Y_2(s) - \hat{\mu}(s)$. Figure 2.3 presents one realization at a time instance for the isotropic field with the mean μ and compares the true field with the fitted values of the field.

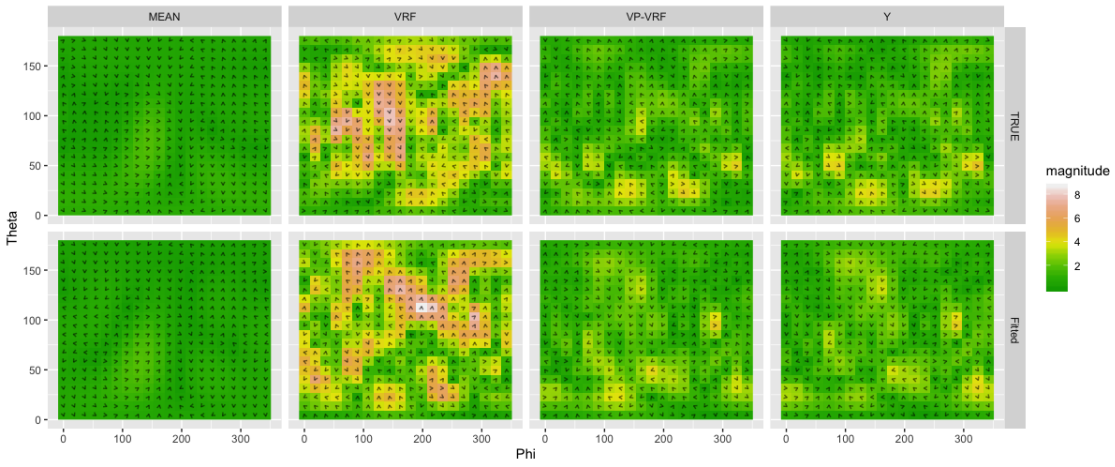


FIGURE 2.4. Anisotropic field with mean field: True VRF refers to the process $y(s)$, and fitted VRF (i.e., the rescaled residual field) refers to the process $(Y(s) - \hat{\mu}(s))/\hat{\varphi}(\theta(s))$. True VR-VRF means $\varphi(\theta(s))(y(s) + \epsilon)$, and fitted VR-VRF is $\hat{\varphi}(\theta)\tilde{y}(s)$, where $\tilde{y}(s)$ is the fitted value of the VRF $y(s)$ using the fitted VRF as the raw data.

In the anisotropic case, we compute an estimate of the scaling factor (by using the integral formula for φ in terms of μ) based on the fitted mean field and divide the mean-subtracted residual field location-wise by the estimated scaling factor in order to obtain the isotropic TVRF, i.e., we use the data $(Y_3(s) - \hat{\mu}(s))/\hat{\varphi}(\theta(s))$. Figure 2.4 shows the comparisons between the true fields and the fitted fields. We observe some discrepancies in terms of the structures of the true and fitted VRF, VR-VRF, and Y fields.

We compare the estimates and evaluate its mean squared errors (MSE, in-sample errors) and predicted mean squared errors (PMSE, i.e., mean squared out-of-sample errors) in Figure 2.5. In the isotropic cases, the estimates are unbiased, and both MSE and PMSE have fairly small values. In the anisotropic cases, estimates for the variance σ_B^2 and σ_C^2 of the random components are underestimated, and estimates for the variance τ^2 of the random noise is overestimated. For the anisotropic process, the scaling function φ is a functional of the mean field μ. The multiplicative nature of the model mean that the estimation errors in $\hat{\mu}$ contributes significantly to the estimation accuracy of the parameters of the isotropic TVRF y. To understand this effect, we consider an additional simulation study where we generate 30 independent realization (corresponding to 30 time instances), rather than 15 in our previous setting. We notice that (see Figures 2.5 and 2.6

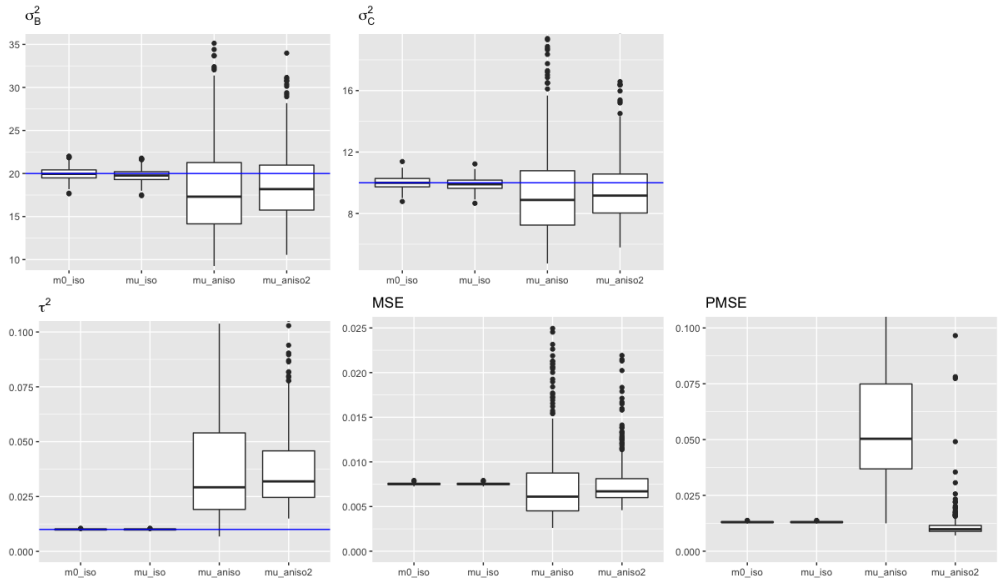


FIGURE 2.5. **m0_iso**: mean zero and isotropic, **mu_iso**: mean μ and isotropic, **mu_aniso**: mean μ and anisotropic, **mu_aniso2**: mean μ and anisotropic ($T = 30$)

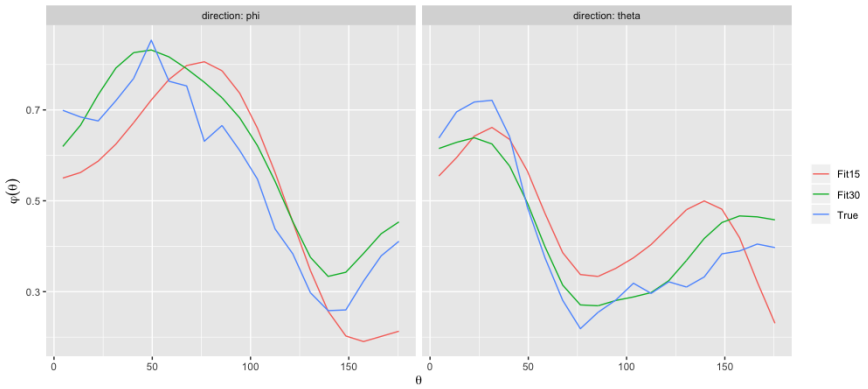


FIGURE 2.6. θ against Anisotropic function and estimated one. Fit15 indicates 15 independent observations of fields, and Fit30 is for 30 independent observations.

for the comparative performance), in this setting (i.e., $T = 30$), due to increased accuracy in the estimation of μ, owing to twice as much as data as before, the accuracies of estimates and predictions are improved. This suggests that, for an accurate estimation and prediction performance, obtaining a more accurate estimate of the mean fields is critical.

Different Mean Magnitude. We also study how the magnitudes of the mean fields affect the estimation and predictions in the anisotropic models, (2.43). We consider four different average magnitudes of the mean fields, so that the average values of the scaling factor φ greater than, less

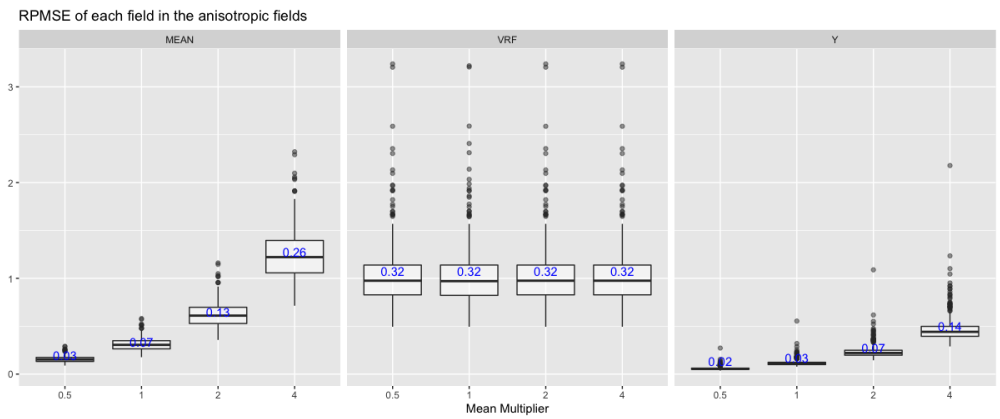


FIGURE 2.7. Different Mean Magnitude against RPMSE in terms of MEAN, VRF(y), Y fields in the anisotropic fields.

than, or close to 1. For comparison, we manipulate the isotropic fields in the identical parameter settings.

The estimates are not affected by the magnitude of the mean fields. The estimates are distributed similarly with similar standard errors across different means. The magnitude effects are found in terms of the in-sample and out-of-sample prediction errors. Whereas the errors for the isotropic fields are roughly constant across the different magnitudes of the mean fields, the errors in the anisotropic fields are proportional to the magnitude of the mean fields. We observe similar levels of errors for the isotropic and anisotropic fields (Figure 2.8) in the case where the magnitude is close to 1 (mu1). In the mu.5 case, i.e., when this magnitude is about 1/2, the error level for the anisotropic fields is lower than the error level in the isotropic field, and, in the cases where the magnitude is greater than 1 (mu2, mu4), the error levels for the anisotropic fields are greater than that for the isotropic field. When we compare the errors in each field such as mean, y, VRF, and Y, we observe that, for the anisotropic field, the errors in VRF tend to increase as the magnitude of mean fields increases; see Figure 2.7. Interestingly, error levels of the y fields are constant across the different means, so the errors from estimating mean fields contribute to the total error levels.

Different Variance Ratio against the Magnitude of Mean field. We have generated TVRFs with the variances,$\sigma_B^2 = 20, \sigma_C^2 = 10$. We now study the case when the fluctuations of the TVRF are more dominant than the mean fields. We investigate both isotropic and anisotropic cases. The anisotropic case mimics our application to geomagnetic modeling. We examine how the ratio

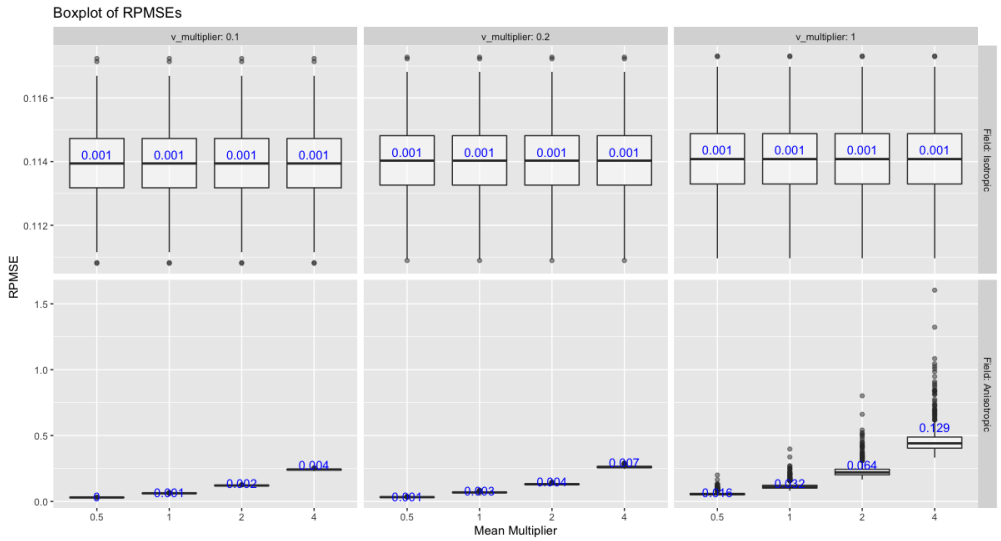


FIGURE 2.8. Mean multiplier against RPMSE: Each column, v_multiplier, is the multiplier of the variance components, and each row is the field.

between the magnitude of mean fields and the variance of random fields influences estimation and predictions.

We have the mean fields with four different magnitudes and add two different sets of variances for comparison with the previous study. The factors multiplying the original values of the variance parameters are 1/10 ($\sigma_B^2 = 2, \sigma_C^2 = 1$) and 1/5 ($\sigma_B^2 = 4, \sigma_C^2 = 2$). The variance τ^2 of the unstructured noise remains unchanged.

The estimates and predictions follow the same pattern as that of the results from the study on varying mean magnitudes. The variances of random components are underestimated in both the isotropic and anisotropic fields. The variances of random errors are overestimated in the anisotropic fields though the variance of random errors is unbiased in the isotropic fields (Figure 2.8). The error levels increase as the magnitudes of the mean fields increase. The size of variance parameters changes the standard errors of estimates, so a larger size of variance has the larger standard errors of estimates. The absolute bias increases as the ratio increases; however, the relative bias remains unchanged.

2.5.3. Iterative Weighted Least Squares. We have found the estimates of random variances are underestimated in the isotropic and anisotropic fields, whereas the variance of the noise is overestimated in anisotropic fields. We conduct iterative weighted least squares for several steps

to inspect whether the overestimation of τ^2 and underestimation of random variances can be mitigated by having a more accurate estimate of the mean that takes into account the correlation structure of the residual (TVRF) field. This is done by formulating the problem of estimation of μ as a weighted least squares problem, where the weight matrix is determined by the inverse of the current fit of the covariance of the random component. This gives rise to an iterative reweighted least squares (RWLS) algorithm for estimation of μ and $\boldsymbol{\theta}$. The updating algorithm of RWLS is as follows. Given kth estimates, $\hat{\boldsymbol{\mu}}^{(k)}, \hat{\boldsymbol{\theta}}^{(k)}$,

$$
\begin{aligned}
W^{(k)} &= M^{-1}(\hat{\theta}^{(k)}, \hat{\mu}^{(k)}), \\
\hat{\mu}^{(k+1)} &= \arg\min_{\mu}(Y-\mu)^T W^{(k)}(Y-\mu), \\
\hat{\theta}^{(k+1)} &= \arg\min_{\theta}(Y-\hat{\mu}^{(k+1)})^T W^{(k)}(Y-\hat{\mu}^{(k+1)}), \\
W^{(k+1)} &= M^{-1}(\hat{\theta}^{(k+1)}, \hat{\mu}^{(k+1)}),
\end{aligned}
$$

where M, the covariance matrix of Y, is defined as $\boldsymbol{\varphi}(\mu)V(\theta)\boldsymbol{\varphi}(\mu)$ in the anisotropic case. In the isotropic case, M is identical to V. The new estimated mean field $\hat{\boldsymbol{\mu}}$ is equal to$Z(Z^TWZ)^{-1}Z^TWY$. We start with the estimates $\hat{\mu}, \hat{\boldsymbol{\theta}}$ from the ordinary least squares and iterate the updating algorithm for three times.

RWLS estimates in the isotropic fields are equivalent to the original estimates (OLS). For anisotropic fields, RWLS improves over the OLS based estimation at the first update, and its prediction errors are reduced; see Figure 2.9. No significant improvement in estimates and prediction was observed from the second update onward.

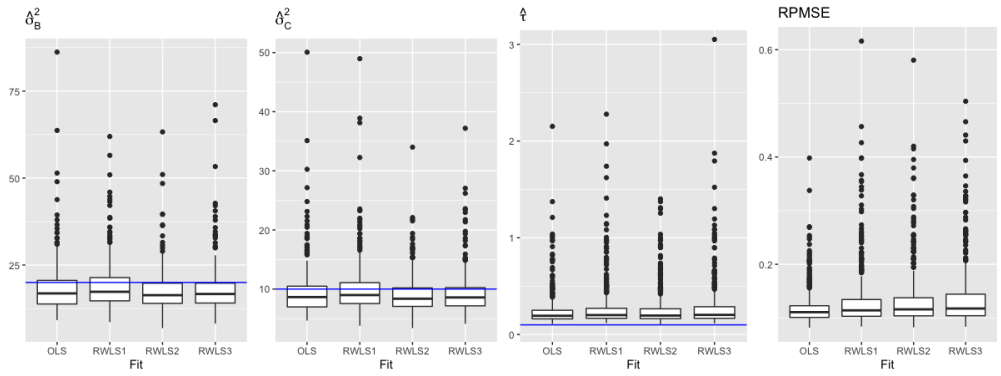


FIGURE 2.9. Results of RWLS in Anisotropic field.

2.5.4. Model Selection. Finding the appropriate model order, L, is formulated as a problem of model selection. Notice that the models for increasing values of L are not "nested" since TVRF model with a larger L does not include the models with a smaller L. We experiment to select the proper L in the simulation data set. Let the true L be 9 and consider it as an unknown parameter. We conduct TVRF modeling with different candidate values, $L = 3, 7, 9, 19$. The results are presented in Figure 2.10.

When a smaller number of VSH basis functions are included the estimates of random variances, σ_B^2, σ_C^2, are underestimated, and their standard errors are larger compared to the true L. The estimates for the variance, τ^2, are overestimates and also have larger standard errors. In this case, due to the under-specification of L, a part of the fluctuations associated with the TVRF, which is not captured by VSH, is clubbed with the unstructured noise. When more VSH bases are included, the variances of random components are underestimated, and the variance of the random noise is similar to the true variance. These results produce the smallest in-sample prediction errors, but the out-of-sample errors are larger that for the true L, which indicates overfitting.

We evaluate model selection criteria such as AIC, AICc, and BIC. The three criteria suggest selecting the maximum order of VSH to be 9, which is the truth, and the RPMSE is also the smallest at $L = 9$.

The results are based on a 20×20 equiangular grid. We conduct the same experiments based on 1600 random locations, and the results agree with those corresponding to the equiangular design. With a sufficient sample size for a certain L, random samples are suitable to estimate the model,

although more computational time is required. An equiangular grid would be a superior choice for modeling TVRFs for computational efficiency if it is applicable.

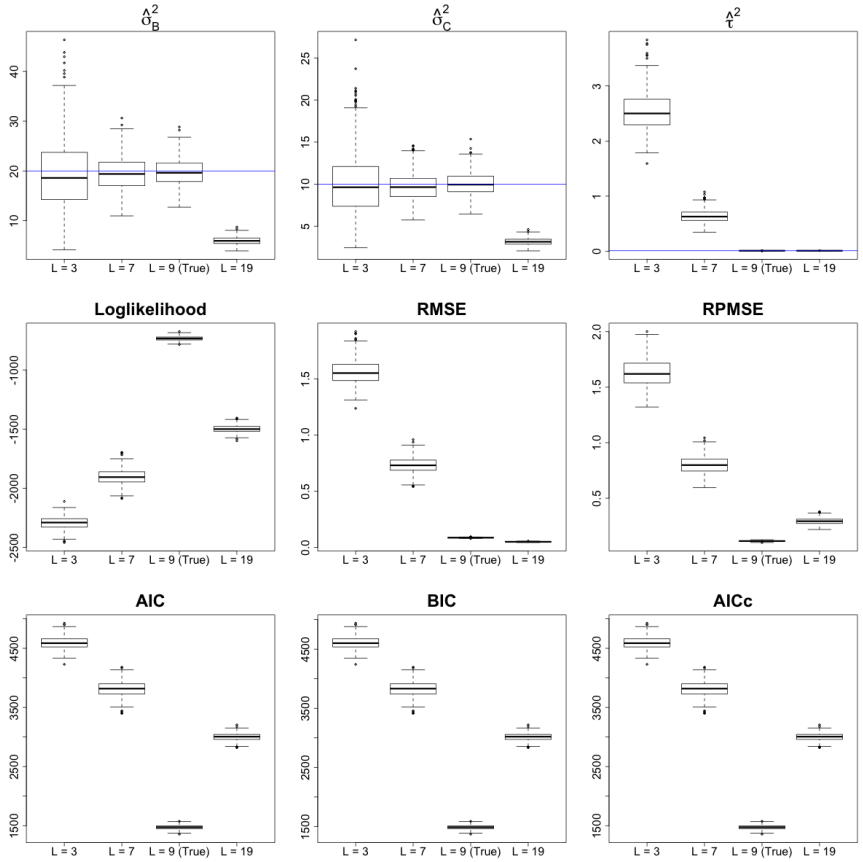


FIGURE 2.10. Predictive Performance: The simulation set is generated with $L = 9$ in 20×20 equiangular grid. Compare estimates with different $L = 3, 7, 9, 19$ at Equiangular grid. The first row is estimation results, and the blue horizontal line indicates the true parameters. Squared root mean squared error (RMSE) is defined as $(\sum_i^n (y_i - \tilde{y}_i)^2/n)^{1/2}$ (in-sample) and squared root predicted mean squared error (RPMSE) is the prediction errors of observations at locations where are not used to estimate parameters (out-of-sample).

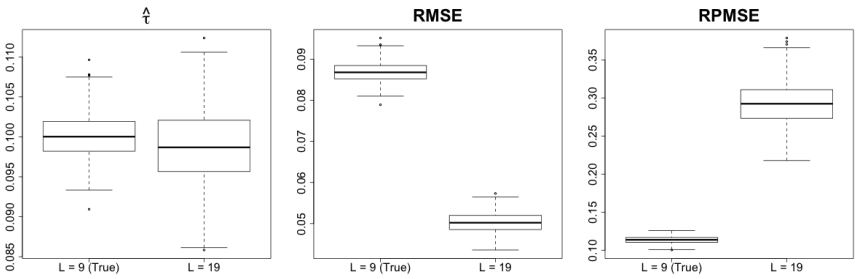


FIGURE 2.11. Comparing $L = 9, 19$ regarding estimated τ, squared root mean squared error and squared root predicted mean squared error.

Model Selection: Anisotropic. We extend the model selection experiment to the anisotropic models (2.43). In this simulation study, we focus on finding the maximum degree, L_μ, of VSH for the mean fields rather than for L_R, the TVRF. Accurate estimates of the mean fields are required for reasonable estimation of the parameters for the anisotropic fields. We generate 15 independent realizations in time for each replicate, which results in 15 independent observations at each location. We set the true L_μ as 3, and the maximum order for the TVRF is set as 9. The variances of random components are two sets, and the variance of random noise is one.

When the smaller maximum degree of VSH is used for fitting the model, the componentwise variances are overestimated, and the corresponding standard errors are larger. When a larger maximum order of VSH for the mean fields is assumed, the componentwise random variances are underestimated, and the noise variance τ^2 is overestimated. The prediction errors from the true L_μ are the smallest among the candidates values.

The model selection criteria recommend a value close to the true L_μ, with some exceptions. The percentages of times either AIC or AICc selects the correct model are approximately 53% each, and that for the BIC is 55%, Figure 2.12. If we select the model based on the smallest PMSE, the percentage of correct selection increases to 69%.

The behaviors of estimates and predictions are rather similar in both the cases of the smaller componentwise variances set and the original componentwise variance set. Model selection performance for the smaller componentwise variances tends to be better. AIC and AICs, select the true L_μ in approximately 63% simulation runs, and BIC selects the true L_μ approximately 91% of

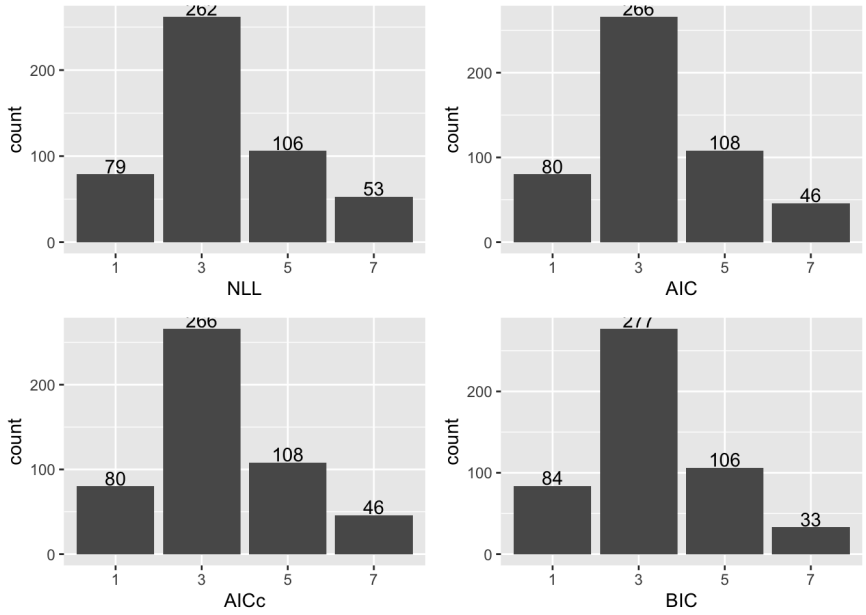


FIGURE 2.12. Histogram of the criteria selections at true L_μ is 3.

times. The PMSE selects the true L_μ in 64% simulation runs. The use of RWLS does not lead to a significant overall improvement over OLS in terms of model selection performance.

Guided by these simulations, we use BIC and the smallest PMSE as the two criteria for model selection in our application.

Smoothness Parameters. We illustrate the impact of the choice of the smoothness parameters (α_B, α_C) on model fitting and prediction in the context of the isotropic model through a simulation study. We assume that smoothness parameters for curl-free and divergence-free fields are equal to α. We estimate $\alpha_B = \alpha_C = \alpha$ by using a model selection criterion by searching over a range of candidate values for α. The results are in Figure 2.13. The variance estimates for random coefficients are dependent on the smoothness parameters since according to our mode $\sigma^2_{\dagger,l} = l^{-\alpha_\dagger}\sigma^2_\dagger,\ \dagger \in \{B, C\}$. No matter what α is chosen, the estimate for the variance of random noise, τ^2, is accurately estimated. MSE and PMSE for the true α are not obviously different from all other $\alpha_\dagger$ candidates. This suggests a relative lack of sensitivity of the fitting results to the specification of α. When the true $\alpha = 2$ is given, the maximum value of the log-likelihood and the minimum of AIC, AICc, and BIC are mostly attained, which suggests that, despite the relative lack of sensitivity, AIC, AICc, and BIC could be suitable criteria for selecting α.

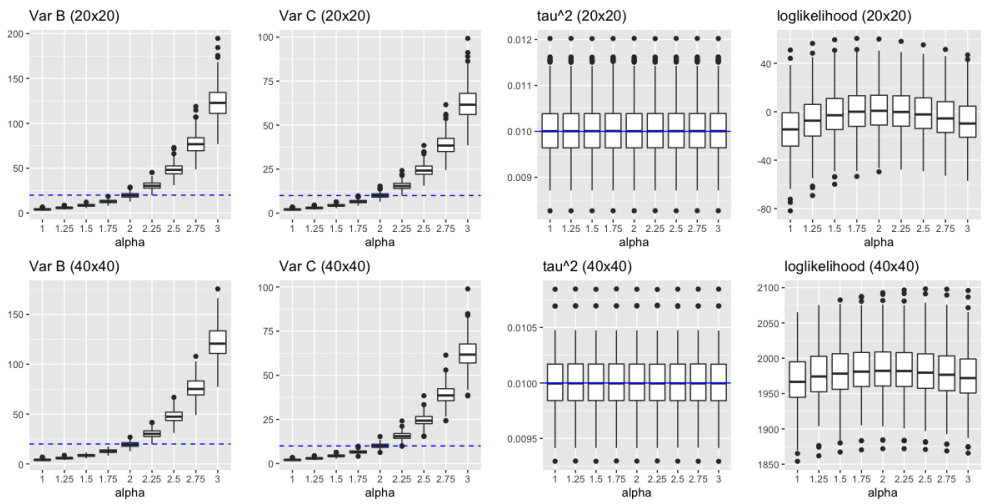


FIGURE 2.13. Experiment of the involving the smoothness parameter $\alpha_\dagger$ for data on (20×20) and (40×40) equiangular grids. The true $\alpha_\dagger$ is 2. Estimates of σ_B^2, σ_C^2 are dependent on $\alpha_\dagger$ candidates, but the estimates of τ^2 are not sensitive to $\alpha_\dagger$.

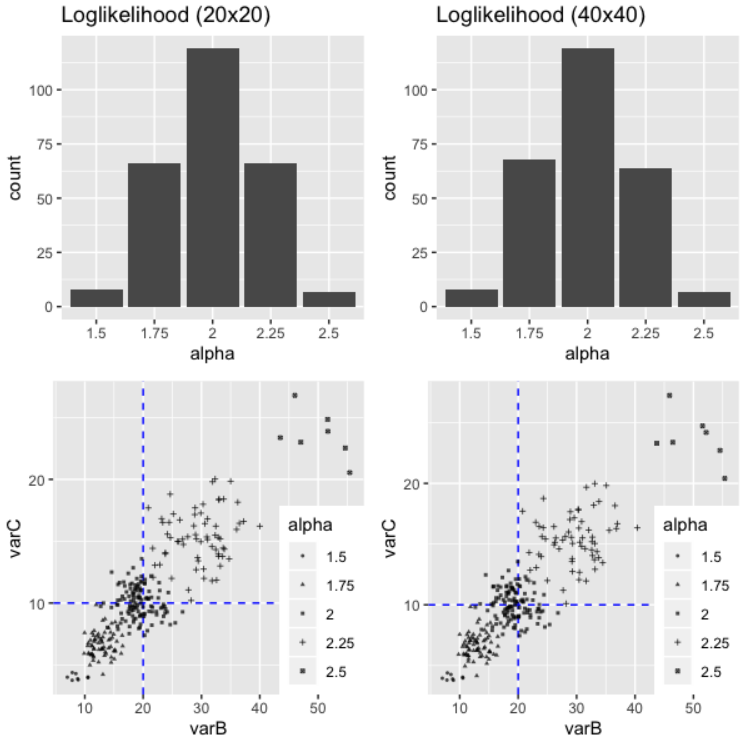


FIGURE 2.14. Selection by AIC: The true $\alpha_\dagger$ is most selected. If $\alpha_\dagger$ is incorrectly chosen, the estimation is far from the true values.

2.5.4.1. *Construct Confidence Sets.* We demonstrate the construction of the parametric bootstrap confidence sets based on the estimates $\hat{\sigma}_B^2, \hat{\sigma}_C^2, \hat{\tau}^2$ in the setting where data comes from a zero mean isotropic TVRF process observed on a 20×20 equiangular grid. We generate parametric bootstrap samples y^* using the MLE of the parameters as the truth and fit the Gaussian TVRF model to find bootstrap estimates $\boldsymbol{\theta}^*$. We have 1500 bootstrap samples ($B = 1500$) of each set of MLEs; see Figure 2.15.

2.5.5. Dependent Case. We also study the case when curl-free fields and divergence-free fields are dependent. As per our modeling scheme, for the TVRF model given by (2.40), we

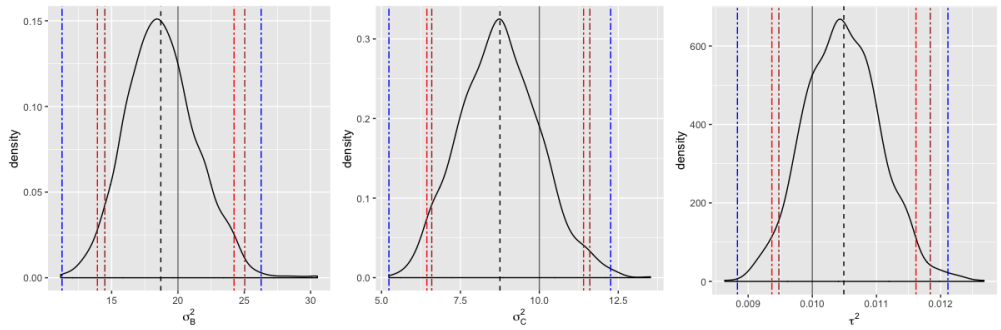


FIGURE 2.15. Bootstrap sample distribution of an independent MLEs. The dotted black line is MLE. The solid line is the parameter. Other Vertical lines indicate the 95% CI methods. (Red: Percentile CI, Brown: Percentile with the bias-corrected bootstrap (BC_a) method, Blue: Asymptotic)

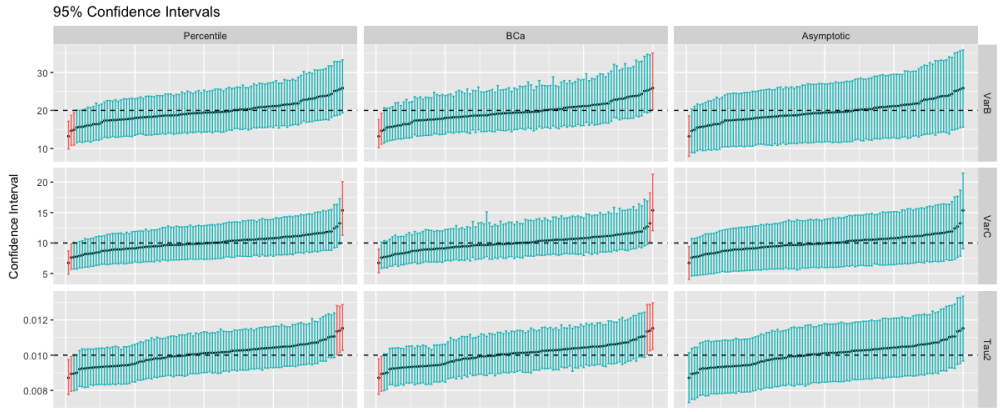


FIGURE 2.16. 100 of 95% confidence intervals for TVRF parameters (Independence case). Each row is for each parameter, and each column for the methods. Green color indicates having true parameters, and red not having the true value. The intervals are arranged by in order of the estimates.

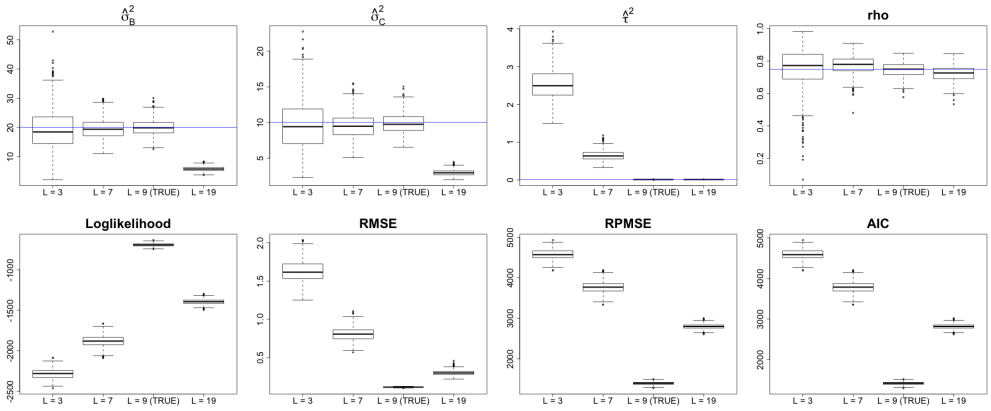


FIGURE 2.17. Cross-Correlation: The blue line indicates the true parameters. The selected estimates are closed to the true value when the true L is given. The L can be selected by AIC.

additionally assign a cross-correlation between the two component fields as $\rho_l = 0.75\mathbf{1}_{(l\leq 9)}$. In terms of the general model description, this corresponds to $\rho = 0.75, L_\rho = 9$ and $\beta = 0$. We select $L_\rho = L$ while estimating the parameters. To be more elaborate, the cross-correlation for each l is identical and exists up to the maximum harmonic frequency (model order) L.

We apply the regression method described in Section 2.4 to obtain initial estimates for σ_B^2, σ_C^2, and ρ. We then use a quasi-Newton method with box constraints to obtain the MLE of $\sigma_B^2, \sigma_C^2, \tau^2$, and ρ. We calculate the corresponding values of the log-likelihood, AIC, RMSE, and RPMSE.

We present the results in Figure 2.17. At the true $L = 9$, all parameters are estimated to be close to the true value, and the corresponding AIC, RMSE, and RPMSE are lowest among all L values. The qualitative behaviors of the estimates of σ_B, σ_C^2, and τ^2 are similar to those in the independent case. The estimated correlation coefficient ρ is seemingly unbiased at smaller or true L, while the standard error of the estimate increases for smaller L. At a larger L, the cross-correlation is underestimated.

Importance of introducing cross-correlation. The cross-correlation helps to understand the TVRF and relationship between curl-free and divergence-free fields. In the equation of VSH, we have $\mathbf{B}^\theta_{l,m} = -\mathbf{C}^\phi_{l,m}, \mathbf{B}^\phi_{l,m} = \mathbf{C}^\theta_{l,m}$. The observed field $\mathbf{Y}$ is expressed as $\mathbf{Y} = \mathbf{Zf} + \boldsymbol{\epsilon}$. If we pull out the

random effect parts for each directions, we derive the following.

$$\begin{aligned}
\mathbf{Z}(s_i)\mathbf{f} &= \sum_{l,m} \left\{ f^B_{l,m}\mathbf{B}_{l,m}(s_i) + f^C_{l,m}\mathbf{C}_{l,m}(s_i) \right\} \\
&= \sum_{l,m} \left\{ f^B_{l,m}\mathbf{B}^{\theta}_{l,m}(s_i)\hat{\boldsymbol{\theta}} + f^B_{l,m}\mathbf{B}^{\phi}_{l,m}(s_i)\hat{\boldsymbol{\phi}} + f^C_{l,m}\mathbf{C}^{\theta}_{l,m}(s_i)\hat{\boldsymbol{\theta}} + f^C_{l,m}\mathbf{C}^{\phi}_{l,m}(s_i)\hat{\boldsymbol{\phi}} \right\} \\
&= \underbrace{\sum_{l,m} f^B_{l,m}\mathbf{B}^{\theta}_{l,m}(s_i)\,\hat{\boldsymbol{\theta}}}_{\text{B_theta}} + \underbrace{\sum_{l,m} f^B_{l,m}\mathbf{B}^{\phi}_{l,m}(s_i)\,\hat{\boldsymbol{\phi}}}_{\text{B_phi}} + \underbrace{\sum_{l,m} f^C_{l,m}\mathbf{C}^{\theta}_{l,m}(s_i)\,\hat{\boldsymbol{\theta}}}_{\text{C_theta}} + \underbrace{\sum_{l,m} f^C_{l,m}\mathbf{C}^{\phi}_{l,m}(s_i)\,\hat{\boldsymbol{\phi}}}_{\text{C_phi}}
\end{aligned}$$

In Figure 2.18, different directional components in the component (B or C) fields exhibit strong linear relationships as the corss-correlations increase (bottom-left: B_theta vs C_phi, top-right: Blm_theta vs Clm_phi plot). Identical directional components in the component fields do not exhibit any effects from the cross-correlations (top-left: B_theta vs C_theta, bottom-right: Blm_phi vs Clm_phi plot).

Estimates of the componentwise variances σ^2_B and σ^2_C are influenced by the cross-correlation, especially when ρ is either -1 or $+1$, and the estimates show positive associations, in Figure 2.19. When the value of $|\rho|$ is less extreme (≤ 0.75 in our simulations), the estimates are very mildly affected.

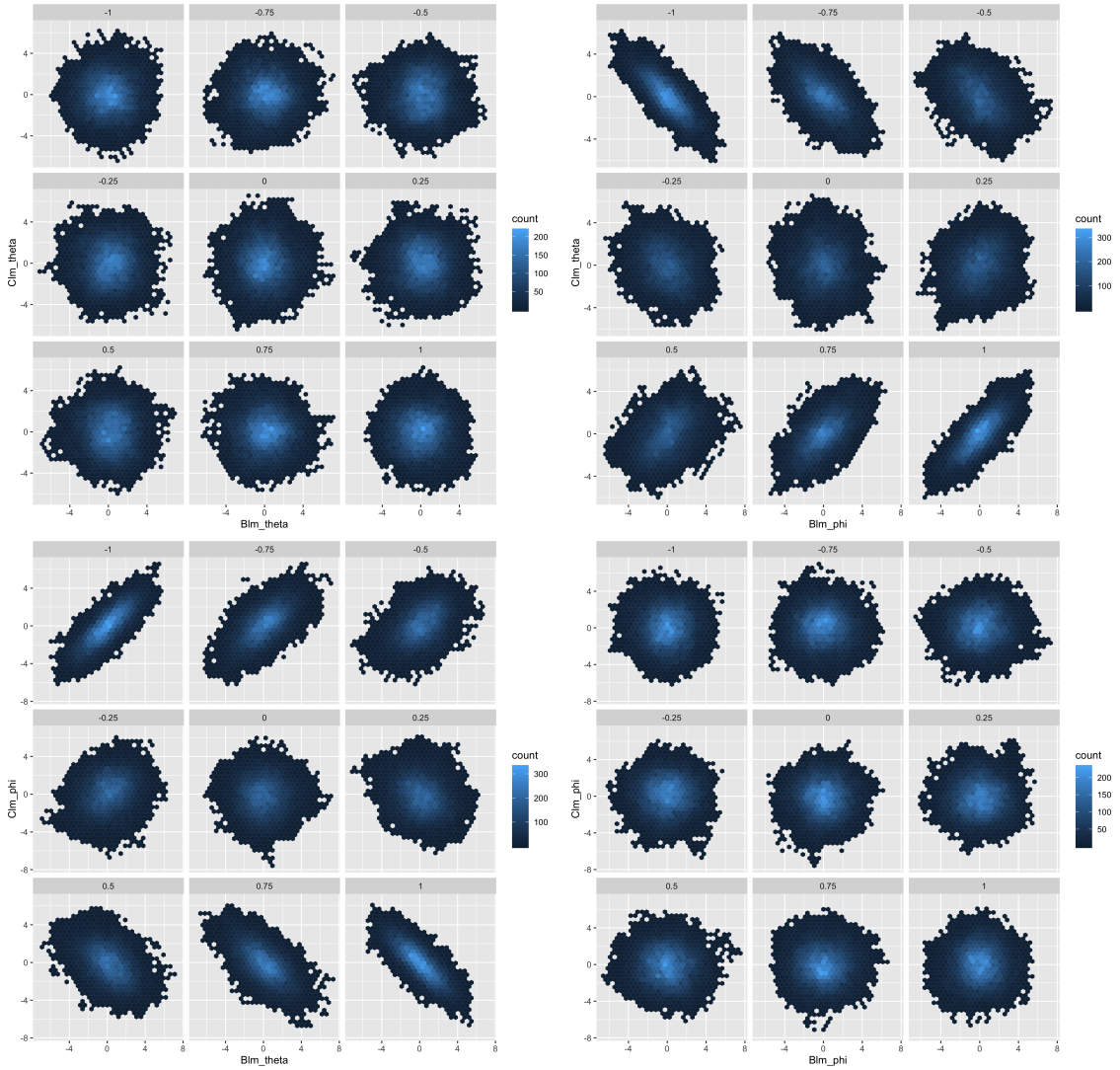


FIGURE 2.18. Plots the different components of the TVRF for varying ρ. Each panel corresponds to one value of the cross-correlation. When we have small or no cross-correlation, the two components do not exhibit a linear relationship. When the cross-correlation is strong, the plots exhibit a linear relationships between the component. Note only 50 replicates are plotted here, $\mathbf{B}_{l,m}$ vs $\mathbf{C}_{l,m}$ at $l = 9, m \geq 0$.

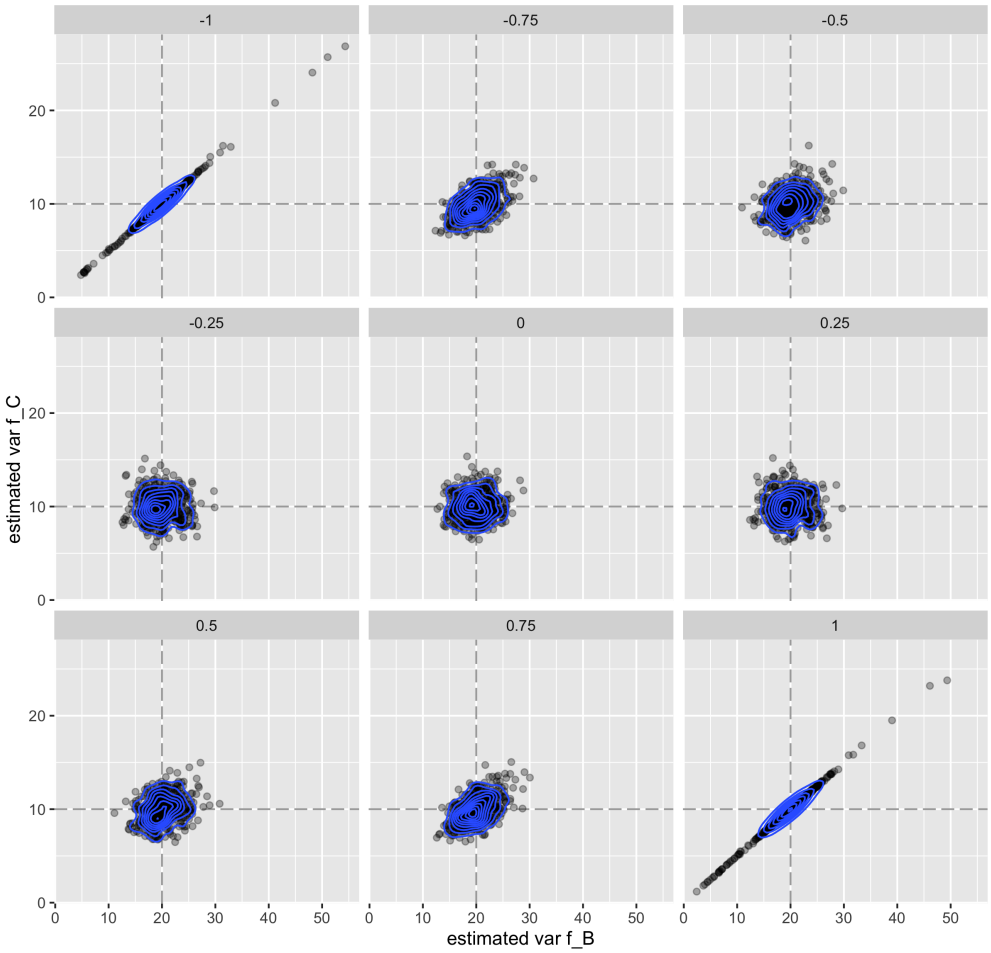


FIGURE 2.19. $\hat{\sigma}_B^2$ vs. $\hat{\sigma}_C^2$: 500 replicates for each true ρ. It exhibits positive linear relationships between the two estimates. The estimates are influenced by the value of ρ. Blue line indicates the contour plot of the joint density of $(\hat{\sigma}_B^2, \hat{\sigma}_C^2)$, and the intersection of the dotted lines indicates the true parameters.

2.6. Analysis of Satellite-Based Measurements on Earth's Magnetic Field

We illustrate the application of the proposed anisotropic TVRF model to Earth's magnetic field modeling based on Ørsted satellite survey data. We summarize current internal geomagnetic field models that are built upon the specification of a potential field, and we show that the proposed model leads to improvement in fitting and prediction performance over some competing methods.

2.6.1. Earth's Magnetic Field Modeling. Researchers have actively studied the problem of modeling Earth's magnetic fields – so-called geomagnetic fields. The geomagnetic field models are used to produce maps, which include extrapolation of the fields that cannot be measured directly, and they form the foundation for the geophysical interpretation of the geomagnetic fields. The measurements of the geomagnetic fields from observatory stations on the ground or from satellites form the different sources of data for modeling this physical process. The sources have different physical characteristics, so researchers determine the contribution of each source in appropriate ways to produce the geomagnetic field models. Depending on the contributions of sources, the geomagnetic fields are distinguished into internal fields and external fields such as ionospheric, magnetospheric, and Earth-induced fields. One of the major scientific challenges is the sophisticated separation of the various fields produced using the sources based on geomagnetic field observations, which requires adequate mathematical representations of the fields [40].

The internal fields, derived from Earth's core and crust, are more than 90% of the geomagnetic measurements and often referred to as the main field. The main magnetic fields ($\mathbf{B}$) can be described by mathematical models as the gradients of a scalar potential V [32].

$$\mathbf{B} = -\nabla V(r, \theta, \phi) \tag{2.44}$$

The potential field V appearing in (2.44) is a scalar harmonic function, that is, it satisfies the Laplace equation $\nabla^2 V = 0$. Hence the resulting vector field B is divergence-free: $\nabla \cdot \mathbf{B} = 0$. Moreover, by *Ampere's Law*, in the absence of an external electrical field, the field $\mathbf{B} = 0$ is also curl-free, i.e., $\nabla \times \mathbf{B} = 0$. In the transitional geomagnetic modeling approach, the scalar potential V is expanded in the SSH basis, and the set of harmonic coefficients of SSH are the parameters of interest (so-called "Gauss coefficients") in the geomagnetic field models. The World Magnetic

Model	POMME-10[35] [36]	WMM [6]	CM5 [43]	OIFM[39]	IGRF-12 [12]	CHAOS-6[11]
Year	2017	2010-2015	2015	2000	2014	2016
Model Type	Combined Model	Main Field	Combined Model	Main Field	Main Field	Main Field
Max SH	133	12	120 (20)	19	13	20
Data	Ørsted, CHAMP, SAC-C	CHAMP, Ørsted, Observatory data	Ørsted, CHAMP, SAC-C, Observatory hourly means	Ørsted	Ørsted, CHAMP, SAC-C, Swarm	Ørsted, CHAMP, SAC-C, Swarm
Model Estimation	Gauss coefficient is given as a truncated Talyor expansion $g(t) = g + tg' + 0.5t^2g''$ Least Squares via eigenvectors and eigenvalues of the normal equation with no regularization	Iterative reweighted least square	Least Squares in the application of iterative Gauss-Newton with linear constraints	Iterative reweighted least square	Iterative reweighted least square	Iterative reweighted least square

TABLE 2.2. Current Available Geomagnetic Models

Model (WMM [6]), International Geomagnetic Reference Field (IGRF-12 [12]) and CHAOS-6 [11] are some of the currently available models for the main geomagnetic field; see Table 2.2 for a comparative summary of these models.

As a further elaboration of this modeling paradigm, the potential field V, defined within an annular shell, is expressed in the spherical coordinate system ((r, θ, ϕ) [32]) as

$$V(r,\theta,\phi) = a\sum_{l=1}^{L_{int}} \left(\frac{a}{r}\right)^{l+1} \sum_{m=0}^{l} (g_l^m \cos(m\phi) + h_l^m \sin(m\phi)) P_l^m(\cos\theta), \qquad r \le a, \tag{2.45}$$

where a is a lower shell (Earth) radius and g_l^m, h_l^m are the Gaussian coefficients. $P_l^m(\cos\theta)$ is the Schmidt quasi-normalized associated Legendre function with order l and frequency m.

The currently available models use equation (2.45) with the assumption that the main fields are curl-free. They have additional extensions or assumptions and different modeling procedures and additional extensions or assumptions, several of which are presented in Table 2.2. Maus et al. [35] [36] developed the model POMME-10 which adopts a sequential approach, conducting data selection, data correction and model fitting sequentially. In this approach, the parameters are separated into groups and sequentially fitted via least squares. Comprehensive Model 5 (CM 5), by Sabaka et al. [43] describes the near-Earth magnetic field. CM 5 parameterizes an empirical inverse model derived from data. The comprehensive inversion method is based on an approach to deriving field models in which the major magnetic field sources are parametrized and co-estimated via least squares. The models are sophisticated and built to describe the physical behaviors of the geomagnetic fields.

In a recent model, CHAOS, [41],[11], $\{g_l^m(t), h_l^m(t)\}$ are time-dependent Gauss coefficients that are modeled using a sixth-order B-spline, i.e., for a 6-th order B-spline basis $\{B_j\}_{j=1}^J$, the coefficients $g_l^m(t)$ is expressed as

$$g_l^m(t) = \sum_{j=1}^{J} g_{l,j}^m B_j(t)$$

and similarly for $\{h_l^m(t)\}$. These researchers use a six-month knots spacing with five-fold repeated knots at the endpoints $t = 1997.1$ and $t = 2016.6$.

2.6.2. Ørsted Satellite. Before we illustrate our modeling applications to the geomagnetic field observations by satellites, we give a descriptive summary of the features of satellite-based measurements and their variability characteristics, which in incorporate in our model for the observations. The satellite data offer a dense coverage of the earth, except near the poles. The observations are collected while the satellites orbit the earth, so it is difficult to isolate the sources of variations in the measurements across, as to whether they are due to the location or the time changes – which is referred to as a time-aliasing effect. The satellites also travel through electrical plasma-filled space, and the existence of electric currents at the satellite altitude implies that, it may not be appropriate to describe the observed field as the gradient of a harmonic potential, as in (2.44) (see [40] for details).

Because of the issues mentioned above, if we are interested in modeling only the horizontal component of the magnetic field, it is better not to impose too stringent a structural constraint (as in (2.44)) on the observed vector field. The TVRF model, and its anisotropic extensions, provide the flexibility as well as interpretability needed to model the fluctuations of such fields. Accordingly, treating the observed data as a spatio-temporal vector field measured at a fixed altitude (the latter is an approximation), we apply the proposed methodology to the Ørsted Satellite data set. The Ørsted satellite was launched on February 23, 1999, and one of its missions is to map the geomagnetic fields [19]. We obtain high precision magnetic field data (Ørsted Prelim Mag-L) [1], such as time, satellite position vectors, Magnetic Field Magnitude, and Magnetic field vectors.

Orbit Trajectory. We take samples from the trajectories of the Ørsted satellite measurements, which allows us to examine and rectify for the time aliasing effect. We define a *trajctory* as an orbit from the north pole back to the north pole. The satellite needs to traverses 101 trajectories to cover the surface of the earth, so as to reach the original position (approximately) and the total time span of these 101 trajectories is approximately a week. We focus our analysis on the 15 such time spans covering the period from January 2001 to April 2001. We observe that each trajectory has its own characteristics our preliminary descriptive statistical analysis suggests that conducting sampling along the trajectories is recommended. We conduct functional data analysis on the geomagnetic observations against longitude, treating data along similar trajectories (observed after a fixed time interval), as functional observations, and verify the time dependency of the observed field. We further investigate this time dependency by examining the auto-correlation coefficients of the functional principal components scores. We observed a mild time dependency, so we model the time dependency by absorbing all the temporal variation in the mean field, by modeling the VSH coefficients of the mean field as smooth time-varying functions (see (2.50)).

2.6.2.1. *Fluctuations.* The random fluctuations of the raw Ørsted satellite data are anisotropic [21]. The fluctuations are caused by only one sensor location on Ørsted satellites to collect vector data, and it varies according to the altitude of the satellite. It is known that the uncertainty of the random fluctuations is related to the magnitudes of the components [20]. We use this information to model the variance of the random fluctuation based on the magnitude of the mean fields. The

[1]The data sets are available to the public at DTU space website. `https://www.space.dtu.dk/english/research/scientific_data_and_models/magnetic_satellites`

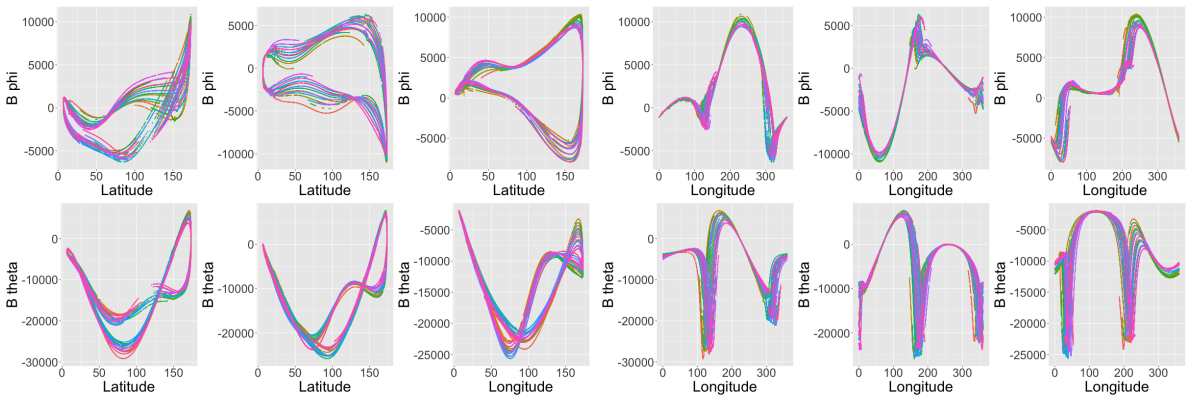


FIGURE 2.20. Plots of latitude (θ) and longitude(ϕ) against each component of $\mathbf{B}$. Each panel indicates a different set of trajectories, and different colors correspond to different time periods.

magnitudes of the mean fields change with latitude (θ) and increase to the south pole from the north pole. The observed relationship between the variance and the mean field can be written as follows:

$$Var(\widetilde{\mathbf{y}}(\theta_i, \phi_i, t)) \propto \int |\boldsymbol{\mu}(\theta_i, \phi, t)|^2 d\phi, \tag{2.46}$$

where $\widetilde{\mathbf{y}}(\theta_i \phi_i, t)$ is the random fluctuation at location θ_i, ϕ_i at time t. Equation (2.46) expresses that the variance is dependent only on the latitude, which indicates the axial symmetry of the components of the vector field.

2.6.3. TVRF Model for Ørsted Satellite Data. We use an extended version of our TVRF model to describe the Ørsted satellite data set. We treat the data as the horizontal component of the magnetic field and treat it as having been observed at a fixed altitude (which is an approximation). This means that the observed field can be treated as a tangential vector field measured along the satellite trajectories. Furthermore, we treat model the divergence-free and curl-free components of the field to be independent (so that $\rho_l \equiv 0$ in our modeling paradigm). We assume that the observed tangential field (i.e., the horizontal component of the magnetic field), is the sum of a non-random and time-varying trend ($\boldsymbol{\mu}$) and anisotropic random stochastic fluctuations ($\widetilde{\boldsymbol{y}}$). We define

the observed geomagnetic vector field $Y(s_i,t)$ at $s_i \in \mathbb{S}^2$ and time t, $i = 1, \cdots n_t, t = 1, \cdots, T$, as

$$Y(s_i,t) = \mu(s_i,t) + \widetilde{y}(s_i,t) \tag{2.47}$$

$$= \mu(s_i,t) + \varphi(s_i,t)\big(y(s_i,t) + \epsilon_i(t)\big), \tag{2.48}$$

where $\mu(s_i,t)$ is a non-random time varying trend and $\widetilde{y}(s_i,t)$ is random stochastic fluctuation at location s_i and time t. As noted from (2.46), the value of the scaling function, or standard deviation profile, $\varphi(s_i,t)$, is related to $\boldsymbol{\mu}(s_i,t)$. Indeed, based on the empirical relationship (2.46), model $\varphi(s,t)$ as $\sqrt{\int |\mu(\theta(s),\phi,t)|^2}$, where $\theta(s)$ denotes the θ coordinate of the location s. This shows that $\varphi(s,t)$ is only a function of the latitude $\theta(s)$, and consequently, the random component of the field $Y(s,t)$ described in (2.48) is assumed to be axially symmetric. Note that, we use the two components of the vector $\boldsymbol{\mu}(s,t)$ separately to describe the scaling functions associated with the two coordinates of the vector field. Thus, the anisotropic random field, $\tilde{\boldsymbol{y}}$, is a noise-corrupted version of the $\boldsymbol{y}$, multiplied (coordinatewise) by the axially symmetric scaling function. We model latter field as an isotropic Gaussian TVRF.

We suggest two steps in the modeling: (1) non-random and time-varying mean field modeling and (2) non-stationary random field component modeling.

Time-Varying Trend $\boldsymbol{\mu(s,t)}$. We perform the linear regression to fit the mean trend after representing the VSH basis coefficients of the mean field in a cubic B-spline basis to cope with the time variation. Thus, we model the tangential vector mean field $\boldsymbol{\mu}(s_i,t)$ as

$$\boldsymbol{\mu}(s_i,t) = \sum_{l=1}^{L_\mu} \sum_{m=-l}^{l} \big\{ h^B_{l,m}(t)\mathbf{B}_{l,m}(s_i) + h^C_{l,m}(t)\mathbf{C}_{l,m}(s_i) \big\} \tag{2.49}$$

$$= \sum_{l=1}^{L_\mu} \sum_{m=-l}^{l} \sum_{j=1}^{J} \big\{ b^B_{l,m,j}\mathcal{B}_j(t)\mathbf{B}_{l,m}(s_i) + b^B_{l,m,j}\mathcal{B}_j(t)\mathbf{C}_{l,m}(s_i) \big\}, \tag{2.50}$$

where $\{\mathcal{B}_j\}_{j=1}^J$ is a collection of cubic B-spline basis functions, with equally spaced knots, covering the time domain under consideration, and L_μ is the maximum degree of VSH. The fixed coefficients $\{h^B_{l,m}(t)\}, \{h^C_{l,m}(t)\}$ are therefore expressed in the spline basis such that $h^B_{l,m}(t) = \sum_{j=1}^J b^B_{l,m,j}\mathcal{B}_j(t), h^C_{l,m}(t) = \sum_{j=1}^J b^C_{l,m,j}\mathcal{B}_j(t)$, where $\{b^B_{l,m,j}\}$ and $\{b^C_{l,m,j}\}$ are unknown model parameters (regression coefficients) to be estimated from the data.

Tangential Vector Random Fields. We obtain observations on the anisotropic random fields after subtracting the fitted values of the time-varying trend from the observations. We use the estimated mean field to obtain estimates of the scaling factor φ. We model the axially symmetric scaling functions for each direction separately as $\varphi^{\theta}(\theta_i, t) = \sqrt{\int |\boldsymbol{\mu}^{\theta}(\theta_i, \phi, t)|^2 d\phi}$ and $\varphi^{\phi}(\theta_i, t) = \sqrt{\int |\boldsymbol{\mu}^{\phi}(\theta_i, \phi, t)|^2 d\phi}$. In the application, in order to compute the estimates of φ^{θ} and φ^{ϕ}, we divide the earth parallelly into 20 equally space latitudinal (θ) bins, and add the squares of the components of $\hat{\boldsymbol{\mu}}(\cdot, t)$ within the bin containing a given θ_i, and integrate over ϕ numerically to approximate the value of $\varphi^{\theta}(\theta_i, t)$ and $\varphi^{\phi}(\theta_i, t)$ for each θ_i.

After subtracting the mean and rescaling the resulting field $\mathbf{y}(s, t)$ is modeled as an (noisy) isotropic TVRF, and further, we assume the fluctuations to be independent in time, so that, for different times t, the realizations $\mathbf{y}(s, t)$ are i.i.d. Thus,

$$\mathbf{y}(s_i, t) = \frac{\mathbf{Y}(s_i, t) - \boldsymbol{\mu}(s_i, t)}{\varphi(\theta_i, t)} = \sum_{l=1}^{L_R} \sum_{m=-l}^{l} [f^B_{l,m}(t)\mathbf{B}_{l,m}(s_i) + f^C_{l,m}(t)\mathbf{C}_{l,m}(s_i)] + \epsilon_i(t), \tag{2.51}$$

where L_R is the model order for the TVRF, and $\epsilon_i(t)$ denotes random noise that are assumed to be i.i.d. both across time and location, and independent across coordinates, with zero mean and variance τ^2. $f^B_{l,m}(t), f^C_{l,m}(t)$ are Gaussian random coefficients with zero mean and variance $\sigma^2_{B,l}, \sigma^2_{C,l}$, which are modeled as $\sigma^2_{\dagger,l} = \sigma^2_{\dagger} l^{-\alpha_{\dagger}}$, for $1 \le l \le L_R$, and $\dagger \in \{B, C\}$. After obtaining the "observed values" of the isotropic TVRF through the equation (2.51), in which we substitude $\boldsymbol{\mu}$ by its least squares estimate, we find the MLE of the parameters $\boldsymbol{\theta} = (\sigma^2_B, \sigma^2_C, \tau^2)$, initially treating L_R, α_B and α_C as given.

Log-Likelihood Including Mean Trend. We consider the full log-likelihood including the mean fields, to select the final model. Let $\mathbf{Y}_t = \mathbf{Y}(\mathbf{s}, t) = \big(Y_t(s_1), Y_t(s_2), \cdots, Y_t(s_{n_t})\big)$ where $i = 1, \cdots, n_t, t = 1, \cdots, T$. The TVRF model for Ørsted data is defined as

$$\mathbf{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varphi}_t(\mathbf{y}_t + \boldsymbol{\epsilon}), \tag{2.52}$$

where $\boldsymbol{\varphi}_t = \mathrm{diag}[\varphi_t(s_1), \cdots, \varphi_t(s_{n_t})]$.

The log-likelihood is

$$\ell(\boldsymbol{\theta}) = -\frac{\sum_{t=1}^{T} n_t}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\Big[\log(|V_t^{\varphi}(\boldsymbol{\theta})|) + \text{trace}\big[V_t^{\varphi}(\boldsymbol{\theta})^{-1}(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})^T\big]\Big] \tag{2.53}$$

where $V_t^{\varphi}(\boldsymbol{\theta}) = \boldsymbol{\varphi}_t[Z_t G Z_t^T + \tau^2 I_{n_t}]\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_t[V_t(\boldsymbol{\theta})]\boldsymbol{\varphi}_t$, where $G = G(\boldsymbol{\theta})$ is as in equation (2.25).

For estimating the random fields, we minimize a simpler version of the log-likelihood (2.54).

$$\begin{aligned}
-2\ell(\boldsymbol{\theta}) &= \ const. \ + \sum_{t=1}^{T}\Big[\log(|V_t^{\varphi}(\boldsymbol{\theta})|) + \text{trace}\big[V_t^{\varphi}(\boldsymbol{\theta})^{-1}(\mathbf{Y}_t - \boldsymbol{\mu}_t)(\mathbf{Y}_t - \boldsymbol{\mu}_t)^T\big]\Big] \\
&= \ const. \ + \sum_{t=1}^{T}\Big[\log(|\boldsymbol{\varphi}_t[V_t(\boldsymbol{\theta})]\boldsymbol{\varphi}_t|) + \text{trace}\big[(\boldsymbol{\varphi}_t[V_t(\boldsymbol{\theta})]\boldsymbol{\varphi}_t))^{-1}(\mathbf{Y}_t - \boldsymbol{\mu}_t)(\mathbf{Y}_t - \boldsymbol{\mu}_t)^T\big]\Big] \\
&= \ const. \ + \sum_{t=1}^{T}\Big[\log(|V_t(\boldsymbol{\theta})|) + \text{trace}\big[V_t(\boldsymbol{\theta})^{-1}\mathbf{y}_t\mathbf{y}_t^T\big]\Big], \qquad (2.54)
\end{aligned}$$

where the constant does not depend on $\boldsymbol{\theta}$. The BLUP of Y are estimated to be

$$\mathbb{E}\big(\mathbf{Y}_t(\mathbf{s}_{new})|\mathbf{Y}_t(\mathbf{s}_{old})\big) = \hat{\boldsymbol{\mu}}_t + \widehat{\boldsymbol{\varphi}}_t\Big[\mathbf{Z}_{t,new}\widehat{G}\mathbf{Z}_{t,old}^T V_{t,old}^{-1}(\hat{\boldsymbol{\theta}})(\mathbf{Y}_t\big(\mathbf{s}_{old}) - \hat{\boldsymbol{\mu}}_t(\mathbf{s}_{old})\big)\Big],$$

where, the right hand side should be interpreted as a plug-in estimate of the BLUP (i.e, the EBLUP), $\boldsymbol{Z}_{t,old}$ is the evaluation matrix of VSH at the observed locations, $\mathbf{Z}_{t,new}$ is an evaluation matrix at the new locations, and $V_{t,old}(\boldsymbol{\theta})$ is the submatrix of $V_t(\boldsymbol{\theta})$ corresponding to the vector of observation locations $\mathbf{s}_{old}$.

2.6.4. Applying our model on the Ørsted Satellite Data. We illustrate the application of the method described above to the Ørsted satellite data. In our parametric modeling, we estimate fixed coefficients of the time-varying mean, variances of the random coefficients and variance of random observational noise for different models. We consider AIC, BIC, AICc, and PMSE to select the final model. We compare our final model with a few competing procedures for fitting the vector field, and the results illustrate that our model has better prediction performance, indicating the usefulness of including the random effects component of the TVRF in model fitting.

We consider 15 time periods of Ørsted satellite data from January 2, 2000, to April 16, 2000. As discussed earlier, each time period corresponds to 101 orbital trajectories. We randomly sample points from each trajectory evenly to cover the whole earth, while ensuring at least a total sample

size of 1600 to satisfy the required sampling rate for the maximum order of spherical harmonics to be 19. The trajectories are not always fully observed, and for each time period, parts of some trajectories are missing. We draw random locations from each trajectory in proportion to the number of the total available observations in the trajectory. This way, we obtain samples of 18 to 24 locations from the trajectories within each time period, so we have 1645 sampling locations for the first time period. The total number of sampled locations for all periods is 24601. We sample another new data set for cross-validation. The out-of-sample data are collected from the same periods of time. Each period has a sample size of at least 400, and the total size of the out-of-sample data set is 6702.

We model the TVRF with the mean fields (2.52). We estimate the mean fields with various candidate models with different specifications of $(L_\mu, \alpha_B, \alpha_C)$, while keeping the TVRF model order $L_R = 19$ fixed for all the cases. We also allow the coefficients to be independent of time (the setting denoted by "ind" in Table 2.3). For the time-dependent models (settings indicated by "dep" in Table 2.3), we use cubic B-splines with intercept and one knot, so that, $J = 5$ in equation (2.50). After we fit the mean fields, we estimate the variance of random coefficients with given smoothness parameters (α_B, α_C). We assume the smoothness parameters are greater or equal to 1.

	L_μ	k	α_B	α_C	$\hat{\sigma}_B^2$	$\hat{\sigma}_C^2$	$\hat{\tau}^2$	RMSE	RPMSE	Time	Field
1	9	498	1	1	3.033×10^{-5}	5.657×10^{-5}	8.761×10^{-4}	280.74	311.24	dep	blm
2	7	318	1	1	5.498×10^{-5}	5.634×10^{-5}	8.733×10^{-4}	278.18	311.61	dep	blm
3	9	102	4	2	4.948×10^{-2}	4.656×10^{-4}	8.857×10^{-4}	285.11	313.01	ind	blm
4	9	201	4	2	4.874×10^{-2}	4.239×10^{-4}	8.873×10^{-4}	285.77	313.15	ind	both
5	9	102	3	2	7.688×10^{-3}	4.647×10^{-4}	8.751×10^{-4}	282.04	313.20	ind	blm
6	9	201	3	2	7.633×10^{-3}	4.248×10^{-4}	8.766×10^{-4}	282.65	313.38	ind	both
7	7	129	3	2	9.294×10^{-3}	4.641×10^{-4}	8.772×10^{-4}	281.24	313.86	ind	both
8	9	993	3	2	1.947×10^{-9}	3.573×10^{-5}	9.322×10^{-4}	304.77	314.36	dep	both
9	9	993	1	1	1.316×10^{-5}	2.086×10^{-5}	8.874×10^{-4}	305.69	315.72	dep	both
10	7	633	1	1	4.074×10^{-5}	2.803×10^{-5}	8.831×10^{-4}	311.47	318.60	dep	both

TABLE 2.3. Results of candidates models are arranged by squares root predict mean squared error (RPMSE). L_μ is the maximum degree of VSH. The field column indicates what fields are used for the estimated mean fields. blm indicates using only a curl-free field, and both mean using both the curl-free and divergence-free field. The time column indicates whether assuming time dependency or independence. The final chosen model is in the top row.

We select the final model (2.55) in which the mean field is expressed in the curl-free fields and is varied over time by cross-validation. We evaluate the MSE and PMSE of each model. The final model has the smallest RPMSE among all candidates. It fulfills our assumption about the non-random and time-varying mean fields; the main field is curl-free fields.

$$\mathbf{Y}(s_i,t) = \underbrace{\sum_{l=1}^{L_\mu=9}\sum_{m=-l}^{l}\{h^B_{l,m}(t)\mathbf{B}_{l,m}(s_i)\}}_{\boldsymbol{\mu}(s_i,t)} + \boldsymbol{\varphi}(\theta_i,t)\underbrace{\left(\sum_{l=1}^{L_R=19}\sum_{m=-l}^{l}\{f^B_{l,m}(t)\mathbf{B}_{l,m}(s_i)+f^C_{l,m}(t)\mathbf{C}_{l,m}(s_i)\}+\epsilon_i\right)}_{y(s_i,t)+\epsilon_{i,t}} \tag{2.55}$$

For comparison, we have used random forest method(RF) [4], and generalized boosted regression model (GBM) [42] to fit the same data set. We have fitted each direction separately against latitude, longitude, and time (considering two univariate models) as the predictor variables. We select the tuning parameters for these methods to produce the smallest prediction error (RMSE). The parameters for RF are the minimum node size and sampling fraction. The parameters for GBM are the maximum depth of each tree, the minimum number of observations in the terminal nodes of the trees, and a fraction of the training set.

	RF	GBM	VSH	VRF
RMSE	1200.1718	351.1564	310.1380	280.7446
PRMSE	1189.4378	392.3918	314.7705	311.2401

TABLE 2.4. Comparison of squared-root mean squared error (RMSE: `sqrt(mean(residuals^2))`) and predicted squared-root squared error(RPMSE: data set not used for estimating parameters.) tables of different models: RF, GBM, VSH: Regression model with curl-free fields VSH $L_\mu = 9$, VRF: TVRF modeling which adds a random effect model on VSH

2.6.5. Confidence Set. Our proposed model enables more accurate and principled uncertainty quantification of various quantities of interest by using the fitted field model. We use approximate sampling distributions of the estimated parameters to obtain confidence sets for the mean field and prediction sets for the mean and TVRF, at any given location on the sphere based on parametric bootstrap sampling. We construct confidence sets for parameters to demonstrate the

bootstrap-based inference procedure. Let $\boldsymbol{\theta} = (L, \alpha_B, \alpha_C, \sigma_B^2, \sigma_C^2, \rho, \tau^2)$, and assume the anisotropic TVRF model

$$Y(s,t) = \mu(s,t) + \varphi(s,t)\big(y(s,t) + \epsilon(s,t)\big).$$

Given $Y(s,t)$, we fit $\hat{\mu}(s,t)$ and $\hat{\varphi}$. We fit the stationary model $\hat{y}(s,t) = \frac{Y(s,t)-\hat{\mu}}{\hat{\varphi}(s,t)}$. Then, we have $\hat{\boldsymbol{\theta}}$. We implement a parametric bootstrap procedure by generating new data as follows:

$$Y^*(s,t) = \hat{\mu}(s,t) + \hat{\varphi}(s,t)\big(y^*(s,t) + \epsilon^*(s,t)\big).$$

$y^*(s,t)$ are independent in t. For each t, $y^*(s,t) \sim GTVRF(\hat{\boldsymbol{\theta}}), \epsilon^*(s,t) \overset{iid}{\sim} N(0, \hat{\tau}^2)$. We obtain $Y^*(s,t)$, and use the fitting procedure described earlier to get $(\hat{\mu}^*(s,t), \hat{\varphi}^*(s,t), \hat{\theta}^*)$ and use the (bootstrap) sampling distribution of these estimates to construct confidence sets for the parameters.

There are two options for sampling for the unstructured noise, $\epsilon^*(s,t)$. The first approach is sampling from $N(0, \hat{\tau}^2)$. The second is from $y(s,t)$ obtain MLE $\hat{\theta}$ and get BLUP $\hat{y}(s,t)$. The residuals are generated from the difference such that $\hat{\epsilon}(s,t) = y(s,t) - \hat{y}(s,t)$. Resampling from $\hat{\epsilon}(s,t)$ may be conducted at given a condition on time t to obtain $\epsilon^*(s,t)$.

We compute B bootstrap estimators from the resamples, and we construct confidence sets from the bootstrapped estimators, $\hat{\theta}^*$ using asymptotic theory or the percentile bootstrap method.

Geomagnetic fields have unknown structures and present a challenge in finding the true model. The potential for model misspecification has negative implications for constructing confidence sets, primarily in terms of coverage of these confidence sets.

2.7. Discussion

We have introduced a TVRF model on the unit sphere. Our model is formulated based on a linear mixed effect model using a VSH representation. The representation can be decomposed into curl-free fields and divergence-free fields based on the Helmholtz-Hodge theorem, so our model endows random effects for the two types of fields. It enables preserving the natural constraints, which are useful for understanding random processes in geophysical and environmental sciences. Our model has a relatively small number of parameters and a simple and interpretable covariance structure.

We examine the performance of our model based on simulated data in different settings. The results show that our methods can consistently estimate parameters and predict appropriately. By model selection criteria like AIC, BIC, and cross-validation, we are able to select the true maximum order of VSH. We also extend our modeling framework to model a certain class of anisotropic fields inspired by that the fluctuations of Ørested satellites are proportional to its mean fields. Because the anisotropic functions are dependent on the mean fields and the presence of the scaling function has a multiplicative effect on the observational noise, the estimation error in the mean field significantly influences our estimates and predictions. The estimates are improved with more observations. Furthermore, one step updating of the estimates using an RWLS procedure is helpful. We may explore different types of anisotropic fields, which depend on the locations but not on the mean fields.

We may study further on how to specify the smoothness parameters. The model selection criteria may not be able to accurately determine the true smoothness parameter, but an incorrect specification of these smoothness parameters influence our estimates of variance components significantly, and it has key roles in terms of understanding the fluctuations of the process.

In applications, we illustrate the modeling of geomagnetic fields from Ørsted satellite data. Our application results suggest that adding the random components improves both fits and predictions, as compared to some standard data fitting schemes that only use fixed effects. We can determine the random effects of each curl-free field and divergence-free field for a geomagnetic field and suggest ways to construct its confidence sets using parametric bootstrap sampling. The final model represents the fields with its physical constraints and quantifies its uncertainties.

Our TVRFs model can be extended to incorporate radial components to manage data on a spherical shell ($\mathbb{S}^2 \times (R_1, R_2)$) for near-earth geophysical application.

$$y(s_i) = \sum_{l=1}^{L} \sum_{m=-l}^{l} \left[f_{l,m}^{R}(r)\mathbf{Y}_{l,m}(s_i) + f_{l,m}^{B}(r)\mathbf{B}_{l,m}(s_i) + f_{l,m}^{C}(r)\mathbf{C}_{l,m}(s_i) \right] + \epsilon_i \tag{2.56}$$

The coefficients are random components and a function of radius, which was not accounted for in the previous TVRFs due to the restriction of the filed to the surface of the unit sphere. We

can further improve computational efficiency to manage large amounts of data by using a discrete Fourier transform if the observation locations are on an equiangular grid.

CHAPTER 3

Inference on Stationary Gaussian Processes on Undirected Graphs

3.1. Introduction

Data with complex high dimensional structures have been observed in many fields recently. In this chapter, we are interested in modeling such data with high dimensional multivariate processes. Very often for such data additional information are available in the form of connectivity of similarity measures.

For example, we consider processes observed on the oceanic surface, where the domain of interest is defined as a subset of a sphere. We assume that sample points are originated from a given deterministic or random manifold. The core structures of the manifold are approximated by a weighted graph constructed from these samples, where each individual point represents a vertex of the graph, and some similarity measure between any pair of vertices represents the weight on the corresponding edge. The phenomenon is illustrated by the characterization of the limits of the graph Laplacian as the Laplace-Beltrami operator on the underlying manifold, under appropriate sampling regimes (See [45] [3] for details).

This chapter proposes a new framework for analyzing data observed on the nodes of an undirected graph using spectral graph theory. We construct Gaussian scalar processes and the associated inferential framework when the domain of observations is an undirected weighted graph in Section 3.2. We study the effects of graph structures and predict the missing values in Section 3.4. We illustrate the model in Gross Domestic Product per capita in the neighborhood graph of counties in Section 3.5.

3.2. Construction of Gaussian Processes on Undirected Graphs

3.2.1. Construction. The stochastic process on a graph is derived from the spectral graph theory [7]. Suppose we have a weighted undirected graph G with N nodes, with symmetric weighted adjacency matrix $\mathbf{W} = (W_{ij})$, where $W_{ij} \geq 0$ for all i, j. Our construction of a stochastic process

is on the spectral decomposition of the graph Laplacian, defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is the diagonal matrix with entries as the degree of the nodes, $D_{ii} = \sum_{j=1}^{N} W_{ij}$.

Consider the spectral decomposition of $\mathbf{L}$

$$\mathbf{L} = \boldsymbol{\chi}\boldsymbol{\Lambda}\boldsymbol{\chi}^T,$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues of $\mathbf{L}$, in increasing order as $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{N-1}$, and $\boldsymbol{\chi} = [\chi_0, \cdots, \chi_{N-1}]$ is the corresponding matrix of orthogonal eigenvectors, with $\chi_0 = N^{-1/2}\mathbf{1}$.

We define a Gaussian process X on the graph G as

$$X = \mu + \sum_{j=0}^{N-1} \sqrt{g(\lambda_j, \boldsymbol{\theta})} Z_j \chi_j, \tag{3.1}$$

where $g(\cdot, \boldsymbol{\theta}) \geq 0$ are spectral density functions parameterized by parameters $\boldsymbol{\theta}$, $\{Z_j\}_{j=0}^{N-1}$ are iid with $N(0,1)$. The process $X = (X(v))_{v \in G}$ has a mean $\mu = (\mu(v))_{v \in G}$ and covariance $\Sigma(\boldsymbol{\theta})$ given by

$$\Sigma_{u,v}(\boldsymbol{\theta}) = \sum_{j=0}^{N-1} g(\lambda_j, \boldsymbol{\theta}) \chi_j(u) \chi_i(v), \qquad u, v \in G \tag{3.2}$$

where $\chi_j(u)$ is the u-th coordinate of the vector x_j. The covariance can be defined as

$$\Sigma(\boldsymbol{\theta}) = \boldsymbol{\chi} g(\boldsymbol{\Lambda}, \boldsymbol{\theta}) \boldsymbol{\chi}^T \tag{3.3}$$

where $g(\boldsymbol{\Lambda}, \boldsymbol{\theta}) = \operatorname{diag}(\lambda_j, \boldsymbol{\theta})_{j=0}^{N-1}$.

We express the observed data Y_t as the process X corrupted by additive random noise:

$$Y_t(v) = X_t(v) + \epsilon_t(v), \qquad v \in G, \quad t = 1, \cdots, T, \tag{3.4}$$

where $\epsilon_t(v)$ are independent and identically normal distributed with a mean of 0 and variance σ^2, and t is observed time.

3.2.2. Spectral Density Function. The statistical model for a Gaussian stationary process on an undirected graph $\mathcal{G}$ is determined by the specification of spectral density function, $g(\lambda, \boldsymbol{\theta})$.

The covariance matrix (3.3) is defined using the eigenvectors of the graph Laplacian $\mathbf{L}$ and the functions $g(\lambda, \boldsymbol{\theta})$, for a given weighted graph $(\mathcal{G}, \mathbf{W})$.

The spectral density function $g(\lambda, \boldsymbol{\theta})$ characterizes the stationary process on the graph. For example, if g is a positive constant function, its covariance matrix becomes an identity matrix or a positive constant multiplied by the identity matrix. In this case, the process has no particular structure but random white noise. If g is a step function like 1 for the case when eigenvalue is 0 or 0 for otherwise, then the covariance matrix will become not a full rank matrix or rank one matrix.

We consider three different spectral density functions.

Model 1 For $j = 0, \ldots, N-1$, set $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$ and

$$g(\lambda_j, \boldsymbol{\theta}) = \theta_1 (j+1)^{-\theta_2} + \theta_4 \lambda_j^{-\theta_3} \mathbf{1}_{\{\lambda_j > 0\}}. \tag{3.5}$$

Here, $\theta_1, \theta_4 > 0$, $\theta_2 > 0$ and $\theta_3 \in [0, 1]$.

Model 2 For $j = 0, \ldots, N-1$, set $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ and

$$g(\lambda_j, \boldsymbol{\theta}) = \frac{\theta_2}{(\theta_1 + \kappa \lambda_j)^{\theta_3}} \ . \tag{3.6}$$

Here, $\theta_1, \theta_2, \theta_3 > 0$, and $\kappa > 0$ is considered as known. This model may be referred to as the *graphical Matérn model* by comparing with the spectral density of a Matérn covariance model.

Model 3 For $j = 0, \ldots, N-1$, set $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ and

$$g(\lambda_j, \boldsymbol{\theta}) = \theta_1 e^{-\theta_2 \lambda_j^{\theta_3}}. \tag{3.7}$$

Here, $\theta_1, \theta_2, \theta_3 > 0$.

Generally, as eigenvalue λ increases, the value of g decreases. As the largest value of g increases, the magnitude of covariance increases, which indicates more fluctuations in the process. The correlation structures are also related to the spectral function g. A process with a smaller range of g values and a larger smoothness parameter shows higher correlations among observations in the connected nodes.

3.2.3. Stationarity. We consider two tests for checking stationarity of observations with respect to the structure of the associated undirected graphs. Schott [44] proposes test statistics as a sum of squared correlations to test the complete independence of random variables having a multivariate normal distribution. The null hypothesis of the test is that there is no correlation between variables, which is written as $H_0 : \rho_{i,j} = 0$ $(i > j)$. The test statistics $t_{n,m}$ is defined in terms of the sample correlation $r_{i,j}$ for $i > j$,

$$t_{n,m} = \sum_{i=2}^{p} \sum_{j=1}^{i-1} r_{i,j}^2 - \frac{p(p-1)}{2(n-1)}, \tag{3.8}$$

where p is the number of variables, n is the sample size. Under the null and assuming that $\lim_{h \mapsto \infty}(p/n) = \gamma \in (0, \infty)$ holds, the test statistic, $t_{n,p}$, converges in distribution to the normal distribution with mean 0 and variance γ^2. The γ^2 is the limit of variance of the test statistic $\lim \mathrm{Var}(t_{n,p}) = \lim \frac{p(p-1)(n-2)}{(n-1)^2(n+1)}$.

Cai and Jiang [5] consider limiting laws of the coherence of a random matrix and suggest testing the covariance structure. Suppose one observes independent and identically distributed p-variate Gaussian variables $Y_1, \cdots, Y_n$ with mean $\mu_{p\times 1}$, covariance matrix $\Sigma_{p\times p}$ and correlation matrix $R_{p\times p} = (r_{ij})$. The null hypothesis to test the covariance structure is:

$$H_0 : \sigma_{i,j} = 0 \qquad \text{for all } |i-j| \geq \delta \tag{3.9}$$

and for a given integer $\delta \geq 1$. Here $\delta = 1$ corresponds to the test of complete independence of the random variables. They define the empirical coherence of the random matrix as follows:

$$L_{n,\delta} = \max_{|i-j|\geq\delta} |r_{i,j}|, \tag{3.10}$$

They show that under the null and some regularity conditions, $(nL_{n,\delta}^2 - 4\log p + \log\log p)$ converges weakly to an extreme value distribution of type I with distribution function $F(y) - \exp\big(-(1/\sqrt{8\pi})e^{-y/2}\big), y \in \mathbb{R}$. Consequently, we reject the null hypothesis at a significant level α if

$$L_{n,\delta}^2 \geq n^{-1}(4\log p - \log\log p - \log(8\pi) - 2\log\log(1-\alpha)^{-1}), \tag{3.11}$$

We use the above two methods to check the stationary in our observations over the undirected graphs, $Y_t(\nu)$. We are interested in checking the stationarity of the covariance structure among the observations on each node under the time independence assumption. We define a matrix $\mathbf{Y}(\nu) = [Y_1, \cdots, Y_t]_{N\times T}$ and $\boldsymbol{S}$ as the sample covariance matrix. Due to the structure of the covariance matrix under the assumption of graph-stationarity, the problem reduces to testing complete independence of the transformed data $\boldsymbol{\chi}^T Y$, where $\boldsymbol{\chi}$ is the matrix of eigenvectors of the graph Laplacian $\mathbf{L}$. In other words, this means testing the null hypothesis that the transformed population covariance matrix $\boldsymbol{\chi}^T \mathrm{Var}(Y)\boldsymbol{\chi}$ is diagonal. We conduct the two tests, by treating $\boldsymbol{\chi}^T \boldsymbol{S}\boldsymbol{\chi}$ as the sample covariance matrix of the observations, compute the test statistics whereby the number of variables (p) is N and the number of sample size (n) is T. We use $\delta = 1$ this purpose, though larger values of δ can be chosen depending on the graph structures. The applications of the tests are illustrated in Section 3.5.

3.3. Maximum Likelihood Estimation

3.3.1. Spatial data. We first consider the maximum likelihood estimation of the model under the spatial setting, i.e., when $T = 1$. We consider the parameter μ to be equal to zero to simplify the analysis. Furthermore, we suppose that we have observations at every single node of the graph $\mathcal{G}$. Then the log-likelihood of the observed data (ignoring additive constants) is given by

$$\log \ell(\boldsymbol{\theta}, \sigma^2) = -\frac{1}{2}\log\left|\Sigma(\boldsymbol{\theta}) + \sigma^2 I_N\right| - \frac{1}{2} Y^T \left(\Sigma(\boldsymbol{\theta}) + \sigma^2 I_N\right)^{-1} Y, \tag{3.12}$$

where $Y = (Y(v))_{v\in\mathcal{G}}$. More generally, if we only observe data at the nodes of a subgraph $\mathcal{S}$, then the corresponding log-likelihood is given by

$$\log \ell_{\mathcal{S}}(\boldsymbol{\theta}, \sigma^2) = -\frac{1}{2}\log\left|\Sigma_{\mathcal{S}}(\boldsymbol{\theta}) + \sigma^2 I_{|\mathcal{S}|}\right| - \frac{1}{2} Y_{\mathcal{S}}^T \left(\Sigma_{\mathcal{S}}(\boldsymbol{\theta}) + \sigma^2 I_{|\mathcal{S}|}\right)^{-1} Y_{\mathcal{S}}, \tag{3.13}$$

where $Y_{\mathcal{S}} = (Y(v))_{v\in\mathcal{S}}$ and

$$\Sigma_{\mathcal{S}}(\boldsymbol{\theta}) = M_{\mathcal{S}} \Sigma(\boldsymbol{\theta}) M_{\mathcal{S}}^T, \tag{3.14}$$

where $M_{\mathcal{S}}$ is a $|\mathcal{S}| \times N$ matrix whose rows are the vectors $\mathbf{e}_v^T$, for $v \in \mathcal{S}$. Here $\mathbf{e}_v$ is a vector in $\mathbb{R}^N$ with the v-th coordinate equal to 1 and the rest equal to zero.

3.3.2. MLE under independent replicates. We consider the scenario where we observe independent realizations of the process. Moreover, we also allow that for different realizations (times), the observations may be obtained on different subgraphs. Specifically, suppose that at time t, for $t = 1, \ldots, T$, we observe the data $\mathbf{Y}_{t,\mathcal{S}_t} = (Y_t(v) : v \in \mathcal{S}_t)$, where $\mathcal{S}_t \subset \mathcal{G}$ is a subgraph with number of nodes equal to $N_t := |\mathcal{S}_t|$. Further, suppose $\bar{N}_T = \min_{1 \le t \le T} N_t$. We assume that $\bar{N}_T \to \infty$ as $T \to \infty$.

Then, the log-likelihood of the observed data is given by,

$$\log \ell_T(\boldsymbol{\theta}, \sigma^2) = -\frac{1}{2} \sum_{t=1}^{T} \log \left| \Sigma_{\mathcal{S}_t}(\boldsymbol{\theta}) + \sigma^2 I_{N_t} \right| - \frac{1}{2} \sum_{t=1}^{T} \mathbf{Y}_{t,\mathcal{S}_t}^T \left(\Sigma_{\mathcal{S}_t}(\boldsymbol{\theta}) + \sigma^2 I_{N_t} \right)^{-1} \mathbf{Y}_{t,\mathcal{S}_t}, \tag{3.15}$$

where $\Sigma_{\mathcal{S}}$ is as in (3.14).

3.3.3. Information matrix. In this section, we compute the information matrix for the process, under the single realization ($T = 1$) for different models in $\mathcal{G}$. Based on the notations introduced earlier, the complete data log-likelihood (3.12) can be expressed as

$$\log \ell(\boldsymbol{\theta}, \sigma^2) = -\frac{1}{2} \sum_{j=0}^{N-1} \log(\sigma^2 + g(\lambda_j, \boldsymbol{\theta})) - \frac{1}{2} \sum_{j=0}^{N-1} \frac{\widetilde{Y}_j^2}{\sigma^2 + g(\lambda_j, \boldsymbol{\theta})}, \tag{3.16}$$

where $\widetilde{Y}_j$ is the $(j+1)$-th coordinate of $\widetilde{Y} = \boldsymbol{\chi}^T Y$. Notice that, under $(\boldsymbol{\theta}, \sigma^2)$, we have $\widetilde{Y} \sim N(0, \sigma^2 I_N + g(\Lambda, \boldsymbol{\theta}))$.

Score function. Let $\boldsymbol{\theta}$ be in an open subset of $\mathbb{R}^d$. Then, for $1 \le k \le d$,

$$\begin{aligned}
&\frac{\partial}{\partial \theta_k} \ell(\boldsymbol{\theta}, \sigma^2 | Y) \\
&= -\frac{1}{2} \sum_{j=0}^{N-1} \frac{1}{\sigma^2 + g(\lambda_j, \boldsymbol{\theta})} \frac{\partial g(\lambda_j, \boldsymbol{\theta})}{\partial \theta_k} + \frac{1}{2} \sum_{j=0}^{N-1} \frac{1}{(\sigma^2 + g(\lambda_j, \boldsymbol{\theta}))^2} \frac{\partial g(\lambda_j, \boldsymbol{\theta})}{\partial \theta_k} \widetilde{Y}_j^2 \\
&= -\frac{1}{2} \mathrm{Trace} \left(\frac{\partial g(\Lambda, \boldsymbol{\theta})}{\partial \theta_k} (\sigma^2 I_N + g(\Lambda, \boldsymbol{\theta}))^{-1} \right) + \frac{1}{2} \widetilde{Y}^T \frac{\partial g(\Lambda, \boldsymbol{\theta})}{\partial \theta_k} (\sigma^2 I_N + g(\Lambda, \boldsymbol{\theta}))^{-2} \widetilde{Y}, \qquad (3.17)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial \sigma^2} \ell(\boldsymbol{\theta}, \sigma^2 | Y) &= -\frac{1}{2} \sum_{j=0}^{N-1} \frac{1}{\sigma^2 + g(\lambda_j, \boldsymbol{\theta})} + \frac{1}{2} \sum_{j=0}^{N-1} \frac{\widetilde{Y}_j^2}{(\sigma^2 + g(\lambda_j, \boldsymbol{\theta}))^2} \\
&= -\frac{1}{2} \mathrm{Trace} \left((\sigma^2 I_N + g(\Lambda, \boldsymbol{\theta}))^{-1} \right) + \frac{1}{2} \widetilde{Y}^T (\sigma^2 I_N + g(\Lambda, \boldsymbol{\theta}))^{-2} \widetilde{Y}. \qquad (3.18)
\end{aligned}$$

Hessian of the log-likelihood. Taking second partial derivatives

$$\begin{aligned}
&\frac{\partial^2}{\partial\theta_k\partial\theta_l}\ell(\boldsymbol{\theta},\sigma^2|Y)\\
&= -\frac{1}{2}\sum_{j=0}^{N-1}\frac{1}{\sigma^2+g(\lambda_j,\boldsymbol{\theta})}\frac{\partial^2 g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k\partial\theta_l}+\frac{1}{2}\sum_{j=0}^{N-1}\frac{1}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^2}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_l}\\
&\quad +\frac{1}{2}\sum_{j=0}^{N-1}\frac{\widetilde{Y}_j^2}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^2}\frac{\partial^2 g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k\partial\theta_l}-\sum_{j=0}^{N-1}\frac{\widetilde{Y}_j^2}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^3}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_l} \qquad (3.19)
\end{aligned}$$

$$\frac{\partial^2}{\partial\theta_k\partial\sigma^2}\ell(\boldsymbol{\theta},\sigma^2|Y) = \frac{1}{2}\sum_{j=0}^{N-1}\frac{1}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^2}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k}-\sum_{j=0}^{N-1}\frac{\widetilde{Y}_j^2}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^3}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k} \qquad (3.20)$$

and

$$\frac{\partial^2}{\partial(\sigma^2)^2}\ell(\boldsymbol{\theta},\sigma^2|Y) = \frac{1}{2}\sum_{j=0}^{N-1}\frac{1}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^2}-\sum_{j=0}^{N-1}\frac{\widetilde{Y}_j^2}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^3}. \qquad (3.21)$$

We denote the Fisher information matrix by $\mathcal{I}(\boldsymbol{\theta},\sigma^2)$ and denote its (k,l)-th element by $\mathcal{I}_{kl}(\boldsymbol{\theta},\sigma^2)$, for $1\le k,l\le d+1$, with first d indexes corresponding to coordinates of $\boldsymbol{\theta}$. Then, for $1\le k,l\le d$,

$$\begin{aligned}
\mathcal{I}_{kl}(\boldsymbol{\theta},\sigma^2) &= -\mathbb{E}_{\boldsymbol{\theta},\sigma^2}\left[\frac{\partial^2}{\partial\theta_k\partial\theta_l}\ell(\boldsymbol{\theta},\sigma^2|Y)\right]\\
&= \frac{1}{2}\sum_{j=0}^{N-1}\frac{1}{\sigma^2+g(\lambda_j,\boldsymbol{\theta})}\frac{\partial^2 g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k\partial\theta_l}-\frac{1}{2}\sum_{j=0}^{N-1}\frac{1}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^2}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_l}\\
&\quad -\frac{1}{2}\sum_{j=0}^{N-1}\frac{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^2}\frac{\partial^2 g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k\partial\theta_l}+\sum_{j=0}^{N-1}\frac{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^3}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_l}\\
&= \frac{1}{2}\sum_{j=0}^{N-1}\frac{1}{(\sigma^2+g(\lambda_j,\boldsymbol{\theta}))^2}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial g(\lambda_j,\boldsymbol{\theta})}{\partial\theta_l}\\
&= \frac{1}{2}\mathrm{Trace}\left[\frac{\partial}{\partial\theta_k}g(\Lambda,\boldsymbol{\theta})\left(\sigma^2 I_N+g(\Lambda,\boldsymbol{\theta})\right)^{-2}\frac{\partial}{\partial\theta_l}g(\Lambda,\boldsymbol{\theta})\right]. \qquad (3.22)
\end{aligned}$$

For $1 \leq k \leq d$,

$$\begin{aligned}
\mathcal{I}_{k,d+1}(\boldsymbol{\theta}, \sigma^2) &= -\mathbb{E}_{\boldsymbol{\theta},\sigma^2}\left[\frac{\partial^2}{\partial\theta_k \partial\sigma^2}\ell(\boldsymbol{\theta}, \sigma^2|Y)\right] \\
&= -\frac{1}{2}\sum_{j=0}^{N-1}\frac{1}{(\sigma^2 + g(\lambda_j, \boldsymbol{\theta}))^2}\frac{\partial g(\lambda_j, \boldsymbol{\theta})}{\partial\theta_k} + \sum_{j=0}^{N-1}\sum_{j=0}^{N-1}\frac{\sigma^2 + g(\lambda_j, \boldsymbol{\theta})}{(\sigma^2 + g(\lambda_j, \boldsymbol{\theta}))^3}\frac{\partial g(\lambda_j, \boldsymbol{\theta})}{\partial\theta_k} \\
&= \frac{1}{2}\sum_{j=0}^{N-1}\sum_{j=0}^{N-1}\frac{1}{(\sigma^2 + g(\lambda_j, \boldsymbol{\theta}))^2}\frac{\partial g(\lambda_j, \boldsymbol{\theta})}{\partial\theta_k} \\
&= \frac{1}{2}\text{Trace}\left[\frac{\partial}{\partial\theta_k}g(\Lambda, \boldsymbol{\theta})\left(\sigma^2 I_N + g(\Lambda, \boldsymbol{\theta})\right)^{-2}\right]. \qquad (3.23)
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathcal{I}_{d+1,d+1}(\boldsymbol{\theta}, \sigma^2) &= -\mathbb{E}_{\boldsymbol{\theta},\sigma^2}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\ell(\boldsymbol{\theta}, \sigma^2|Y)\right] \\
&= -\frac{1}{2}\sum_{j=0}^{N-1}\frac{1}{(\sigma^2 + g(\lambda_j, \boldsymbol{\theta}))^2} + \sum_{j=0}^{N-1}\frac{(\sigma^2 + g(\lambda_j, \boldsymbol{\theta}))}{(\sigma^2 + g(\lambda_j, \boldsymbol{\theta}))^3} \\
&= \frac{1}{2}\sum_{j=0}^{N-1}\sum_{j=0}^{N-1}\frac{1}{(\sigma^2 + g(\lambda_j, \boldsymbol{\theta}))^2} \\
&= \frac{1}{2}\text{Trace}\left[\left(\sigma^2 I_N + g(\Lambda, \boldsymbol{\theta})\right)^{-2}\right]. \qquad (3.24)
\end{aligned}$$

Notice that, the information matrix depends only on the values of $g(\lambda_j, \boldsymbol{\theta})$ and $\frac{\partial}{\partial\theta_k}g(\lambda_j, \boldsymbol{\theta})$, for $k = 1, \ldots, d$, as functions of $\boldsymbol{\theta}$.

3.3.4. Gradient computation for specific models. In this subsection, we summarize the gradients of $g(\lambda_j, \boldsymbol{\theta})$ for each of the three models specified in Section 3.2.2.

3.3.4.1. *Model 1.*

$$\frac{\partial}{\partial\theta_1}g(\lambda_j, \boldsymbol{\theta}) = (j+1)^{-\theta_2}, \qquad (3.25)$$

$$\frac{\partial}{\partial\theta_2}g(\lambda_j, \boldsymbol{\theta}) = -\theta_1 \log(j+1)(j+1)^{-\theta_2}, \qquad (3.26)$$

$$\frac{\partial}{\partial\theta_3}g(\lambda_j, \boldsymbol{\theta}) = -\theta_4 \log\lambda_j \lambda_j^{-\theta_3}\mathbf{1}_{(\lambda_j>0)} \text{ and} \qquad (3.27)$$

$$\frac{\partial}{\partial\theta_4}g(\lambda_j, \boldsymbol{\theta}) = \lambda_j^{-\theta_3}\mathbf{1}_{(\lambda_j>0)}. \qquad (3.28)$$

3.3.4.2. *Model 2.*

$$\frac{\partial}{\partial\theta_1}g(\lambda_j,\boldsymbol{\theta}) = -\frac{\theta_1\theta_3}{(\theta_1+\kappa\lambda_j)^{\theta_3+1}}, \tag{3.29}$$

$$\frac{\partial}{\partial\theta_2}g(\lambda_j,\boldsymbol{\theta}) = \frac{1}{(\theta_1+\kappa\lambda_j)^{\theta_3}} \text{ and} \tag{3.30}$$

$$\frac{\partial}{\partial\theta_3}g(\lambda_j,\boldsymbol{\theta}) = -\frac{\theta_2\log(\theta_1+\kappa\lambda_j)}{(\theta_1+\kappa\lambda_j)^{\theta_3}}. \tag{3.31}$$

3.3.4.3. *Model 3.*

$$\frac{\partial}{\partial\theta_1}g(\lambda_j,\boldsymbol{\theta}) = \exp\left(-\theta_2\lambda_j^{\theta_3}\right), \tag{3.32}$$

$$\frac{\partial}{\partial\theta_2}g(\lambda_j,\boldsymbol{\theta}) = -\theta_1\lambda_j^{\theta_3}\exp\left(-\theta_2\lambda_j^{\theta_3}\right) \text{ and} \tag{3.33}$$

$$\frac{\partial}{\partial\theta_3}g(\lambda_j,\boldsymbol{\theta}) = -\theta_1\theta_2\lambda_j^{\theta_3}\log\lambda_j\exp\left(-\theta_2\lambda_j^{\theta_3}\right). \tag{3.34}$$

3.3.5. Heuristic asymptotic analysis. Given data $(Y(v))_{v\in\mathcal{S}}$, the maximum likelihood estimation involves maximizing $\log\ell_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)$ with respect to $(\boldsymbol{\theta},\sigma^2)$. Below, we conduct a heuristic asymptotic analysis to understand the behavior of the MLE.

Let $\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2) = \Sigma_{\mathcal{S}}(\boldsymbol{\theta}) + \sigma^2 I_{|\mathcal{S}|}$. Let $(\boldsymbol{\theta}_0,\sigma_0^2)$ denote the true parameters for the generative model. Then, by using Gaussianity, we can express

$$Y_{\mathcal{S}} = \left(\Gamma_{\mathcal{S}}(\boldsymbol{\theta}_0,\sigma_0^2)\right)^{1/2}\mathbf{Z}_{\mathcal{S}},$$

where $\mathbf{Z}_{\mathcal{S}} \sim N(0, I_{|\mathcal{S}|})$. Thus, when $N_{\mathcal{S}} := |\mathcal{S}|$ is large, and *subject to some regularity conditions*, by using Gaussian concentration, we can write

$$\frac{1}{N_{\mathcal{S}}}\log\ell_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2) = -\frac{1}{2N_{\mathcal{S}}}\text{Trace}\left[\left(\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^{-1}\Gamma_{\mathcal{S}}(\boldsymbol{\theta}_0,\sigma_0^2)\right] - \frac{1}{2N_{\mathcal{S}}}\log\left|\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right| + O_P(\sqrt{\log N/N_{\mathcal{S}}}), \tag{3.35}$$

where the probabilistic bound (O_P-term) holds uniformly over a compact subset of the parameter values $(\boldsymbol{\theta},\sigma^2)$. Consequently, the difference $\frac{1}{N_{\mathcal{S}}}\log\ell_{\mathcal{S}}(\boldsymbol{\theta}_0,\sigma_0^2) - \frac{1}{N_{\mathcal{S}}}\log\ell_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)$ can be expressed as

$$
\begin{aligned}
&\frac{1}{N_{\mathcal{S}}}\log \ell_{\mathcal{S}}(\boldsymbol{\theta}_0,\sigma_0^2) - \frac{1}{N_{\mathcal{S}}}\log \ell_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2) \\
&= -\frac{1}{2N_{\mathcal{S}}}\text{Trace}\left[I_{|\mathcal{S}|} - \left(\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^{-1}\Gamma_{\mathcal{S}}(\boldsymbol{\theta}_0,\sigma_0^2)\right] - \frac{1}{2N_{\mathcal{S}}}\log\left|\left(\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^{-1}\Gamma_{\mathcal{S}}(\boldsymbol{\theta}_0,\sigma_0^2)\right| + O_P(\sqrt{\log N/N_{\mathcal{S}}}) \\
&= -\frac{1}{2N_{\mathcal{S}}}\text{Trace}\left[I_{|\mathcal{S}|} - \left(\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^{-1/2}\Gamma_{\mathcal{S}}(\boldsymbol{\theta}_0,\sigma_0^2)\left(\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^{-1/2}\right] \\
&\quad - \frac{1}{2N_{\mathcal{S}}}\log\left|I_{\mathcal{S}} - \left[I_{\mathcal{S}} - \left(\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^{-1/2}\Gamma_{\mathcal{S}}(\boldsymbol{\theta}_0,\sigma_0^2)\left(\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^{-1/2}\right]\right| + O_P(\sqrt{\log N/N_{\mathcal{S}}}) \\
&= -\frac{1}{2N_{\mathcal{S}}}\text{Trace}\left[I_{|\mathcal{S}|} - \Delta_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right] - \frac{1}{2N_{\mathcal{S}}}\log\left|I_{\mathcal{S}} - \left[I_{|\mathcal{S}|} - \Delta_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right]\right| + O_P(\sqrt{\log N/N_{\mathcal{S}}}),
\end{aligned}
\tag{3.36}
$$

where

$$
\Delta_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2) = \left(\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^{-1/2}\Gamma_{\mathcal{S}}(\boldsymbol{\theta}_0,\sigma_0^2)\left(\Gamma_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^{-1/2}.
$$

Now suppose that $(\boldsymbol{\theta},\sigma^2)$ is in a small neighborhood of $(\boldsymbol{\theta}_0,\sigma_0^2)$, so that

$$
\|I_{|\mathcal{S}|} - \Delta_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\| = o(1) \tag{3.37}
$$

as $N, N_{\mathcal{S}} \to \infty$, where $\|\cdot\|$ denotes the operator norm.

Then, by expanding the log-determinant term in (3.36) in Taylor series, we have the following approximation

$$
\begin{aligned}
&\frac{1}{N_{\mathcal{S}}}\log \ell_{\mathcal{S}}(\boldsymbol{\theta}_0,\sigma_0^2) - \frac{1}{N_{\mathcal{S}}}\log \ell_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2) \\
&= \frac{1}{2N_{\mathcal{S}}}\text{Trace}\left[\left(I_{|\mathcal{S}|} - \Delta_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^2\right] + R_N(\boldsymbol{\theta},\sigma^2) + O_P(\sqrt{\log N/N_{\mathcal{S}}}),
\end{aligned}
\tag{3.38}
$$

where the residual term $R_N(\boldsymbol{\theta},\sigma^2)$ satisfies

$$
R_N(\boldsymbol{\theta},\sigma^2) = O\left(\|I_{|\mathcal{S}|} - \Delta_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\|\frac{1}{2N_{\mathcal{S}}}\text{Trace}\left[\left(I_{|\mathcal{S}|} - \Delta_{\mathcal{S}}(\boldsymbol{\theta},\sigma^2)\right)^2\right]\right)
$$

.

Now, we analyze the behavior of the dominant term in (3.38) when $\mathcal{S} = \mathcal{G}$, (so that $N_{\mathcal{S}} = N$), that is when we observe data at all the nodes of the graph. In this case, the expressions can be derived analytically that enable us to understand the asymptotic behavior of the MLE of $(\boldsymbol{\theta},\sigma^2)$.

To understand this, notice that, when $\mathcal{S} = \mathcal{G}$, the matrix $\Delta_{\mathcal{S}}(\boldsymbol{\theta}, \sigma^2) \equiv \Delta(\boldsymbol{\theta}, \sigma^2)$ simplifies to

$$\Delta_{\mathcal{S}}(\boldsymbol{\theta}, \sigma^2) = \Delta(\boldsymbol{\theta}, \sigma^2) = \boldsymbol{\chi} \left(g(\Lambda, \boldsymbol{\theta}_0) + \sigma_0^2 I_N \right) \left(g(\Lambda, \boldsymbol{\theta}) + \sigma^2 I_N \right)^{-1} \boldsymbol{\chi}^T.$$

Hence, when $\mathcal{S} = \mathcal{G}$,

$$\begin{aligned} \frac{1}{N_{\mathcal{S}}} \text{Trace} \left[\left(I_{|\mathcal{S}|} - \Delta_{\mathcal{S}}(\boldsymbol{\theta}, \sigma^2) \right)^2 \right] &= \frac{1}{N} \text{Trace} \left[\left(I_N - \Delta(\boldsymbol{\theta}, \sigma^2) \right)^2 \right] \\ &= \frac{1}{N} \sum_{j=0}^{N-1} \left(1 - \frac{g(\lambda_j, \boldsymbol{\theta}_0) + \sigma_0^2}{g(\lambda_j, \boldsymbol{\theta}) + \sigma^2} \right)^2 \\ &= \frac{1}{N} \sum_{j=0}^{N-1} \frac{1}{(g(\lambda_j, \boldsymbol{\theta}) + \sigma^2)^2} \left(g(\lambda_j, \boldsymbol{\theta}) - g(\lambda_j, \boldsymbol{\theta}_0) + \sigma^2 - \sigma_0^2 \right)^2. \end{aligned} \tag{3.39}$$

Again using Gaussian concentration bounds, we can deduce that, for $(\boldsymbol{\theta}, \sigma^2)$ in a compact neighborhood of $(\boldsymbol{\theta}_0, \sigma_0^2)$,

$$\begin{aligned} \frac{1}{T\bar{N}_T} \log \ell_T(\boldsymbol{\theta}, \sigma^2) &= -\frac{1}{2T\bar{N}_T} \sum_{t=1}^{T} \text{Trace} \left[\left(\Gamma_{\mathcal{S}_t}(\boldsymbol{\theta}, \sigma^2) \right)^{-1} \Gamma_{\mathcal{S}_t}(\boldsymbol{\theta}_0, \sigma_0^2) \right] \\ &\quad - \frac{1}{2T\bar{N}_T} \sum_{t=1}^{T} \log \left| \Gamma_{\mathcal{S}_t}(\boldsymbol{\theta}, \sigma^2) \right| + O_P \left(\sqrt{\frac{\log(TN)}{T\bar{N}_T}} \right). \end{aligned} \tag{3.40}$$

3.4. Simulation Studies

We study the characteristics of our graphical Matérn model on an undirected graph, from various scenarios in the simulation studies. We explore how graph structures affect our modeling results. We investigate relationships between parameters of the graphical Matérn model. We demonstrate modeling in cases where we have partial observations on the graph and evaluate the model. We point out that the `optim()` function in `R` is unstable for estimating our model, and hence we use grid search, in spite of its high computational cost, for computing the MLE. We have carried out some numerical comparisons between these approaches.

3.4.1. Graph Structure. We explore how the constructions of graphs affect the estimation of a graphical Matérn modeling. We construct graphs with different node configurations, numbers of nodes, and their connectivity.

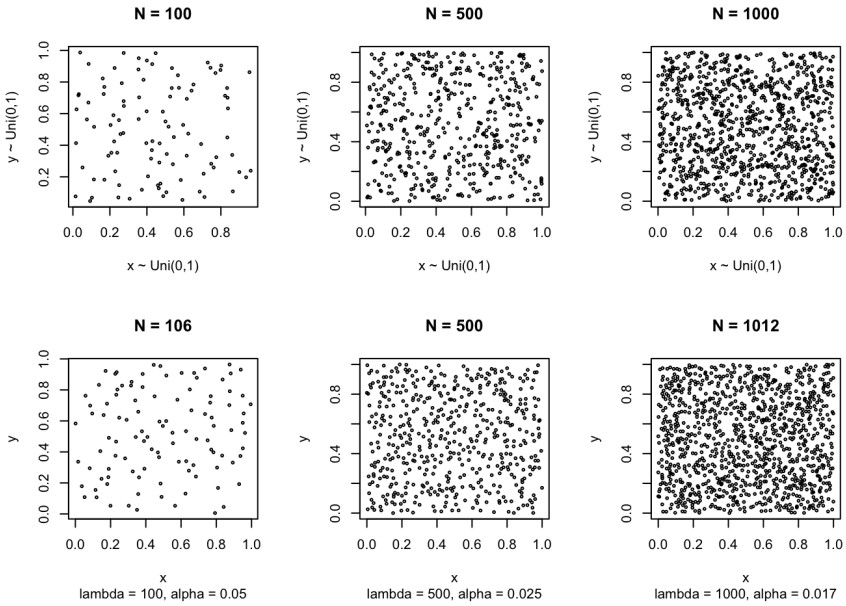


FIGURE 3.1. Node sampling by uniform distribution and Gaussian DPP. The nodes generated by DPP are more evenly spread compared to the nodes by Uniform.

Node Configurations. We use two methods to choose locations of nodes in a unit square: uniform Poisson point process (uniform) or Gaussian determinantal point process (DPP) [33]. In the uniform setting, the coordinate of a node location (x, y) is independently sampled from the unit uniform distribution. We consider a DPP on a unit square, and it is specified by a Gaussian kernel $C_0(x_i) = \rho \exp(-\|x_i/\alpha\|^2),\ x_i \in (0,1) \times (0,1)$, where ρ is an intensity parameter and α is a scale parameter, $\alpha = 1/\sqrt{\pi\rho}$. DPP enforces a degree of repulsion among the nodes (points). Nodes from DPP are more evenly distributed and have fewer open spaces in the domain than nodes with a uniform distribution; see Figure 3.1. We consider 100, 500, and 1000 of node sizes in the uniform configuration. We use $\alpha = 0.05, 0.025, 0.017$ to generate DPP configurations which correspond to ρ equal to 100, 500, and 1000.

Node Connectivity. We create geometric graphs of the node configurations and assign different connectivity among nodes. The weight matrix, $\mathbf{W} = ((W_{ij}))$, of geometric graphs is defined as follows.

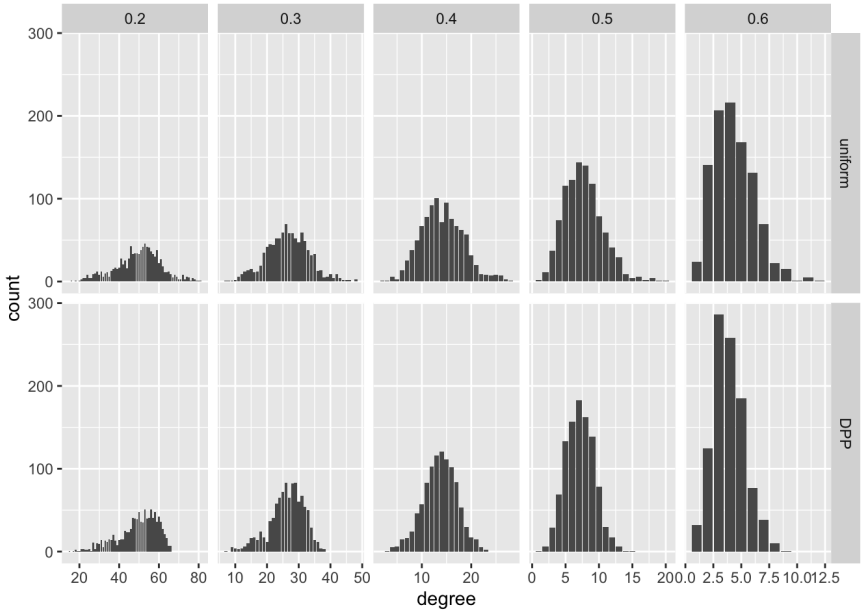


FIGURE 3.2. Degree distributions of the geometric graph with given β and node configurations: Histogram of degree when the number of nodes is 1000. Rows corresponding to distribution of nodes and columns for corresponding β.

For some $\tau > 0$,

$$W_{ij} = c \exp\left(- \frac{\|x_i - x_j\|^2}{\tau_N^2} \right) \tag{3.41}$$

where $c > 0$, and (x_i, x_j) are node coordinates uniformly drawn from a unit square: $(0,1) \times (0,1)$ or from DPP. We define $\tau_N^2 = \tau^2 N^\beta$, where $0 \leq \beta \leq 1$. τ_N^2, which prevents the infill situation in which all nodes are connected as the number of nodes increases, $N \to \infty$. In the experiments, we explore β from 0.2 to 0.6 which gives different edges in the identical node configuration and set $c = 1, \tau^2 = 0.1$ for all graphs. We discretize (W_{ij}) into 0 or 1.

Here β determines the degree of connectivity among the nodes in geometric graphs and influences the degrees of the graphs. The ranges of degrees decrease as β increases, indicating more sparse graphs; see Figure 3.2. The degree distribution at $\beta = 0.4, 0.5$ with DPP nodes are nearly symmetrical, and the degree distribution of uniform nodes are nearly symmetrical at $\beta = 0.3, 0.4$. The distribution with a small $\beta = 0.2$ is left-skewed and, at a larger $\beta = 0.6$, is right-skewed. Degrees of the graphs from DPP are spread in smaller ranges than those from the uniform distribution.

Note that β is closely related to the behavior of eigenvalues of the graph Laplacian because β determines the number of degrees on nodes. We observe the eigenvalues behaving upon the different β and node locations; see Figure 3.3. The largest (first) eigenvalue represents the number of edges on the most connected nodes. The first eigenvalue is bigger in smaller β, and at the same β, the first eigenvalue of uniform node configurations is bigger than that of DPP. The flat slope of the eigenvalue curve indicates that some nodes have a similar number of edges. The slope gets flattered as β increases, and the decreasing rate of eigenvalues is smoother in DPP than in uniform. The number of zero eigenvalues indicates the number of connected graphs, and graphs with larger β include more connected graphs. These are from the fact that nodes in graphs with smaller β tend to have more connections to other nodes and more edges. Moreover, the node configuration of DPP produces more sparse graphs than that of uniform.

We use the spatial information from eigenvalues and eigenvectors of the graph Laplacian and define the covariance structures based on them. In the case of the uniform node configuration at the smallest β, nodes tend to cluster instead spread evenly, and its information is concentrated on particular nodes. Its first eigenvalue is the largest in Figure 3.3, and the corresponding eigenvector is the most fluctuated. We consider such as a difficult environment for modeling, and we expect certain node connectivity to offer a better environment for our model. Next, we explore the quality of our parameter estimates.

Simulation Setting. We generate 500 replicates of the stationary processes with graphical Matérn covariance function (3.6) with $\boldsymbol{\theta} = (\theta_1 = 1, \theta_2 = 1, \theta_3 = 0.25)$ and $\kappa = 0.5$ on the geometric graphs. The node configurations are generated from Uniform and DPP and the connectivity of nodes is defined from different β such as $0.2, 0.3, 0.4, 0.5$, and 0.6. We perform optimization using the Nelder-Mead method to find MLEs, $\hat{\boldsymbol{\theta}}$, for the models with known κ. We optimize two parameters with the other parameter being given, and one initial point is used for the optimization with all 500 simulated data set. The results are reported in Figures 3.5 and 3.6.

We find that the estimates become more accurate for a larger node size with a suitable β. The scatter plots, Figure 3.5, illustrate bivariate distributions of $\hat{\theta}_2$ and $\hat{\theta}_3$ when θ_1 is known at each β and node configurations. When $\beta = 0.5$, the estimates are mostly concentrated at the true parameters. The degree distribution of $\beta = 0.5$ is symmetric in Figure 3.2, and its eigenvalues are

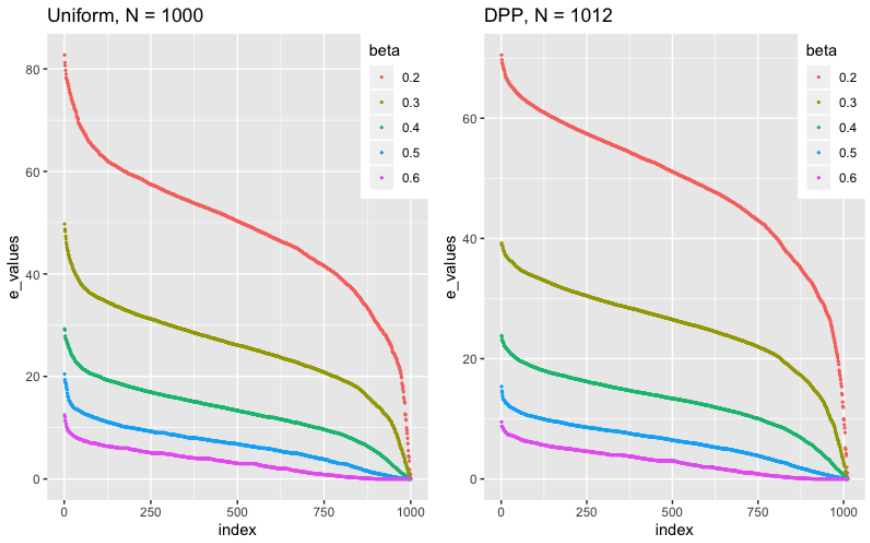


FIGURE 3.3. Eigenvalue plots of the graph Laplacian from uniform (N=1000) and DPP (N=1012)

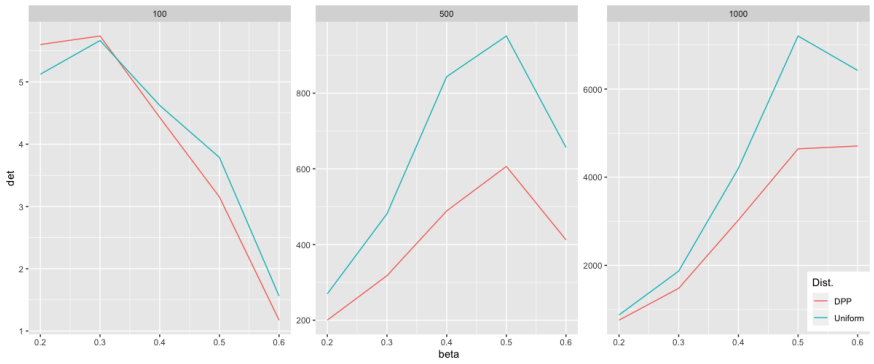


FIGURE 3.4. Determinant of Information Matrix (Subsection 3.3.3) against β per node configuration and size. The determinants are maximized at $\beta = 0.5$ for $N = 500, 1000$ and $\beta = 0.3$ for $N = 100$.

smoothly decreased in Figure 3.3. The determinant of information matrix is maximized at $\beta = 0.5$ when the node size is 500 or 1000, Figure 3.4. It suggests that graphs, having symmetric degree distribution and smoothly decreasing, offer a better environment to identify more precise estimates.

The range parameter, θ_1, is more challenging to be well estimated in any settings; see Figure 3.6. Most estimates of θ_1 are on the boundary of a given possible range, and they are rarely at the true value. Note that θ_1 influences on the estimates of θ_2, θ_3, and when we fix θ_1, the estimates are

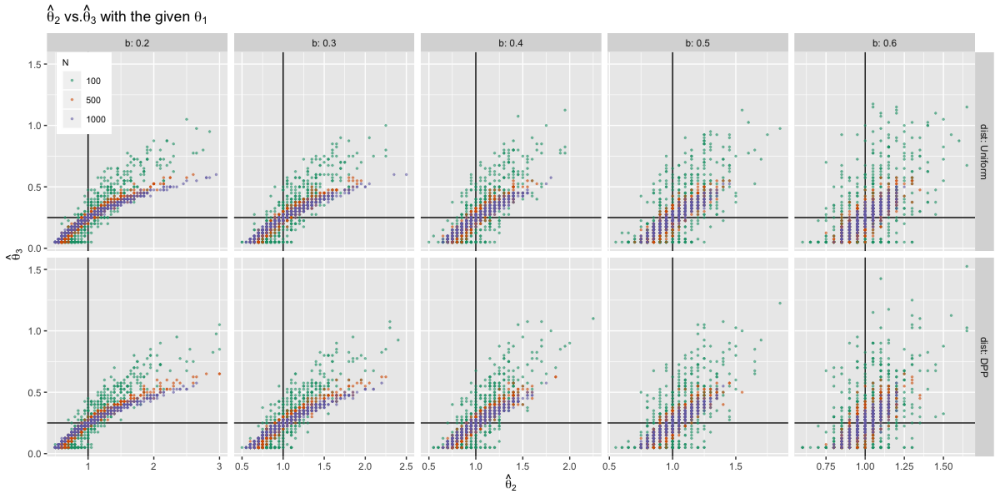


FIGURE 3.5. MLEs of θ_2 and θ_3 with given true θ_1. The reference line denotes the true parameters. N is the number of nodes, and b indicates β. As β increases, the range of MLEs decreases.

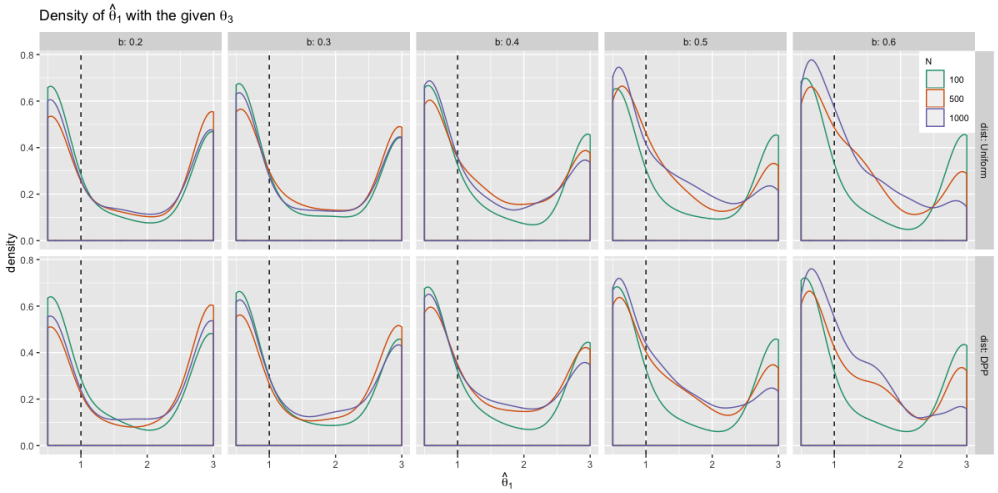


FIGURE 3.6. Marginal density plot of MLE for θ_1 given θ_3. The reference line indicates the true parameter value.

reasonable in Figure 3.7. In our modeling here, we avoid estimating θ_1 directly. We offer discrete fixed values set for θ_1 and estimate the other two parameters for our modeling.

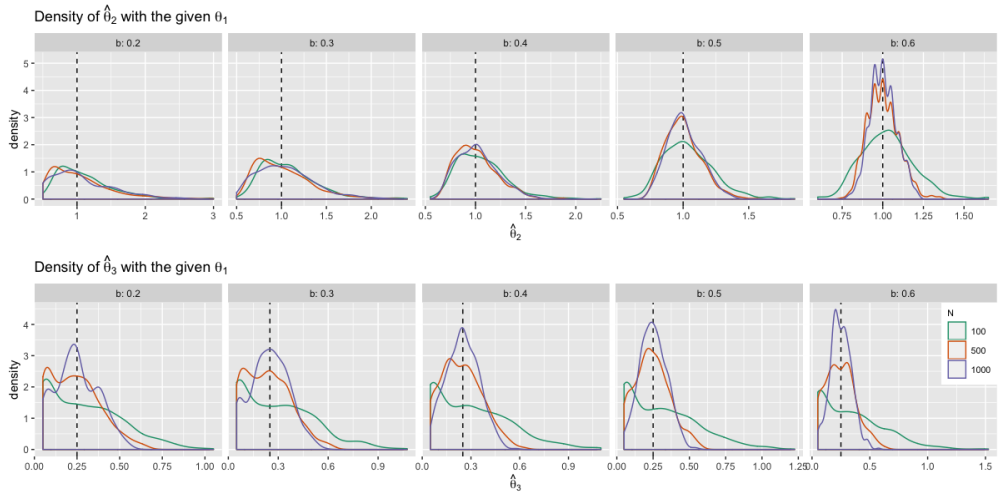


FIGURE 3.7. Density of MLE of θ_2, θ_3: Each row is for θ_2, θ_3. The columns are corresponded to β. The dotted gray line indicates the true parameter.

We find that β in geometric graphs is critical to obtain quality estimates. First, β influences on the behavior of eigenvalues. If $\beta = 0$, the same as $\tau_N^2 = \tau^2$, the eigenvalues are smoothly distributed, so it is hard to estimate model parameters. If $\beta = 1$, which indicates a few connected nodes, this case does not properly convey its spatial structure into the model. Second, β affects on the degree distributions in Figure 3.2. As β increases, the range of degree distributions decreases. We observed skewness on the degree distribution of larger or smaller β. Figure 3.7 illustrates the distributions of estimated θ_2, θ_3, and the MLEs are closed to the true parameter at $0.4 \leq \beta \leq 0.6$. When $\beta = 0.5$, it produces the most accurate estimation quality and exhibits symmetrical degree distribution overall. The determinant of the information matrix is maximized at $\beta = 0.5$; see Figure 3.4.

3.4.2. Partial observations on a graph. We illustrate situations where we have partial observations on an undirected graph. We investigate how our models perform differently upon the structure of observation, the graph size, and the node connectivity. We explore the following settings, such as different rate (γ) of observed values, missing locations, graph size (N), and the node connectivity (β).

- Rate of observation: $\gamma = 0.2, 0.5, 0.8$
- Location of observation: Random or Block

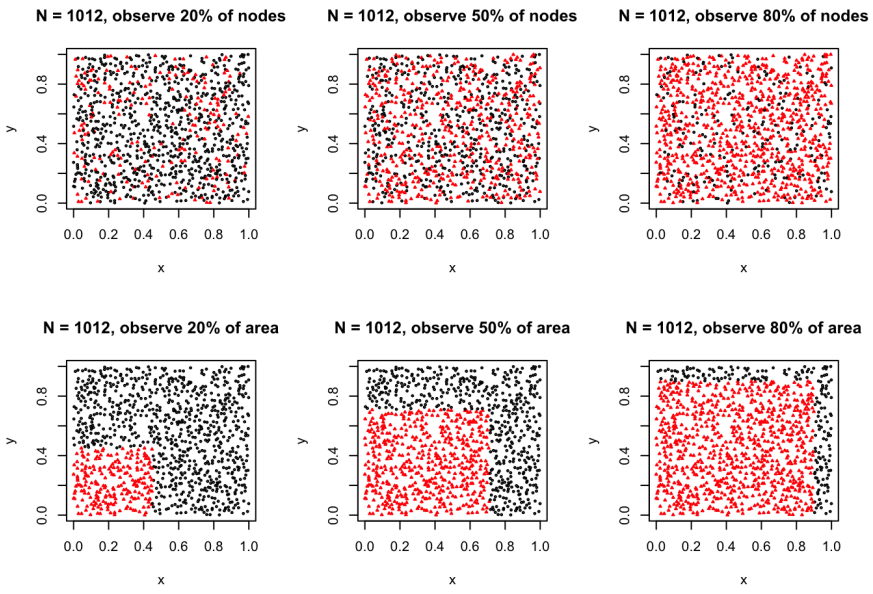


FIGURE 3.8. The node locations are generated from DPP.

- Node connectivity: $\beta = 0.25, 0.5, 0.75$
- Graph Size $N = 500, 1012$

We generate 100 Gaussian processes with the graphical Matérn covariance function (3.6). We demonstrate two parameter sets such as $(\theta_1 = 1, \theta_2 = 1, \theta_3 = 0.25, \sigma^2 = 0.01)$ and $(\theta_1 = 1, \theta_2 = 1, \theta_3 = 0.5, \sigma^2 = 0.01)$ with $\kappa = 0.5$. We estimate the parameters of our model with the given θ_1 and κ using observed values (3.13) and predict the unobserved values. We explore how the qualities of estimation and prediction vary over the graph structures and observations.

We confirm that the sparsity of the graph connection (β) affects the estimation and predictions. For a denser graph with many edges and smaller β, the estimates are centered at the true parameters and have larger standard errors; see Figure 3.9. For a sparse graph, the estimates tend to be biased with smaller standard errors. In-sample errors, (MSE: $\frac{1}{|\mathcal{S}|}(Y_{\mathcal{S}} - \widetilde{Y}_{\mathcal{S}})^2$) and out-of-sample errors (PMSE) increase as β increases; see Figure 3.10. The prediction errors increase in more sparse graphs. To model the processes in graphs properly, we need to have certain connectivity levels among the nodes.

The rate of observations (γ) does not influence the estimates of the models and predictions generally. A larger area (rate) or larger graph size in terms of node sizes produces smaller estimates'

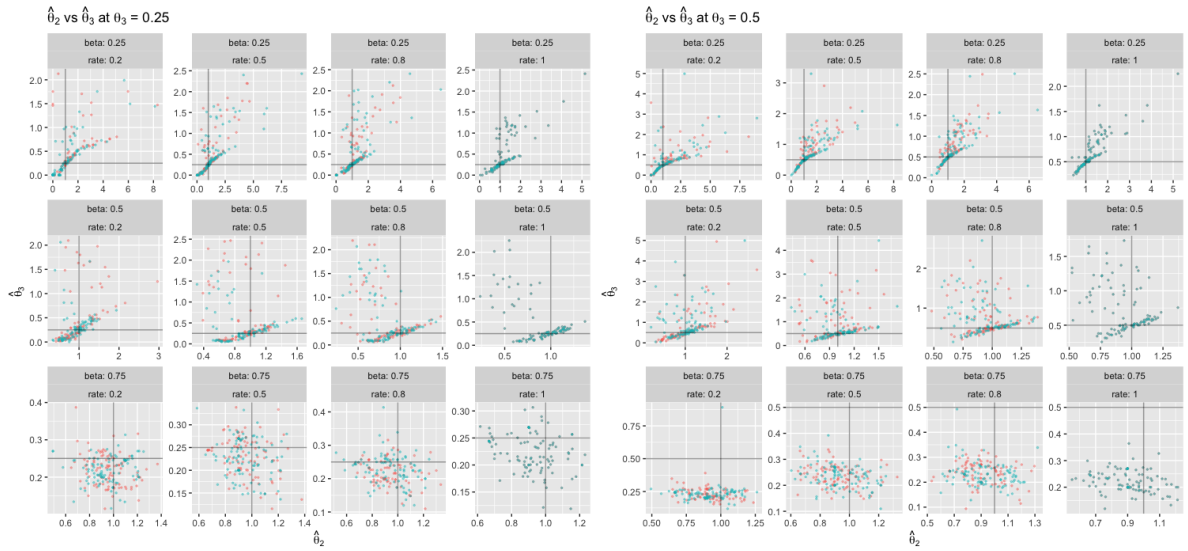


FIGURE 3.9. Scatter plots between $\hat{\theta}_2$ and $\hat{\theta}_3$. Red is the estimate from random observed locations, and green is from block locations. The reference lines indicate the true parameters.

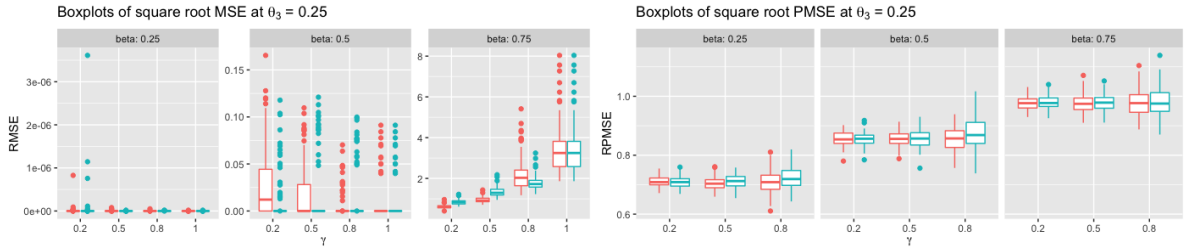


FIGURE 3.10. Boxplots of RMSE and RPMSE against the rates of observations: Each panel represents the β of the graphs. The prediction error is dependent on the sparsity of graph (β) rather than the rate (γ).

standard errors. In-sample errors are different over the rate in Figure 3.10, and out-of-sample errors are constant levels over the rate. The errors are affected by the sparsity of the graph rather than the rate of observations. At the low rate of observations and $\beta = 0.5$, the estimates of random observed locations give higher errors than the block locations. At the low rate of observations and $\beta = 0.75$, the estimates of random locations present smaller errors. The quality of estimates and predictions are from the graph structures which convey sufficient information of processes.

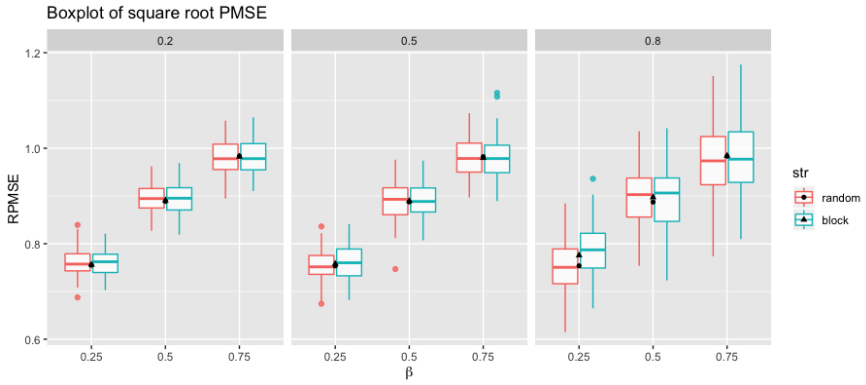


FIGURE 3.11. Boxplots of RPMSE against β and its theoretical conditional standard deviations. The theoretical conditional standard deviation is black dot for random and is black triangle for block.

We compare RMSE and RPMSE of our experiments to the theoretical conditional variance (3.42). In Figure 3.11, the errors increases as β increases no matter how accurate the estimates are:

$$\Gamma_{BB|AA} = \frac{1}{|B|}\mathrm{tr}\Big(\Gamma_{BB} - \Gamma_{BA}\Gamma_{AA}^{-1}\Gamma_{AB}\Big), \tag{3.42}$$

where A refers the observed nodes and values, and B for the unobserved. Block locations are better for the prediction of in-sample in Figure 3.11. There is no significant difference between block and random locations in out-of-sample, and RPMSE is more related to the sparsity of the graph.

Estimating the variance (σ^2) of random fluctuations is challenging in any given situations; see Figure 3.12. Its estimates have larger standard errors at denser graphs. The estimates tend to be biased at a sparse graph, and they likely converged at the initial point of the optimization. When we have bigger smoothness parameter $\theta_3 = 0.5$, the standard errors of the estimates tend to be smaller than at $\theta_3 = 0.25$. However, if we have a larger smoothness parameter θ_3 for a better estimate of σ^2, θ_3 becomes trickier to be estimated. We will further explore the relationships among parameters of the graphical Matérn model and the variance of random fluctuation in Section 3.4.3.

In the experiments, we observe that the estimation and prediction are difficult in the situation with large missing values (small γ) in a denser graph (small β). The connectivity of the graph affects the estimation and prediction of the models significantly.

3.4.3. Explore Variables in graphical Matérn Model. We perform experimental studies to explore connections between parameters of the graphical Matérn model and the variance of

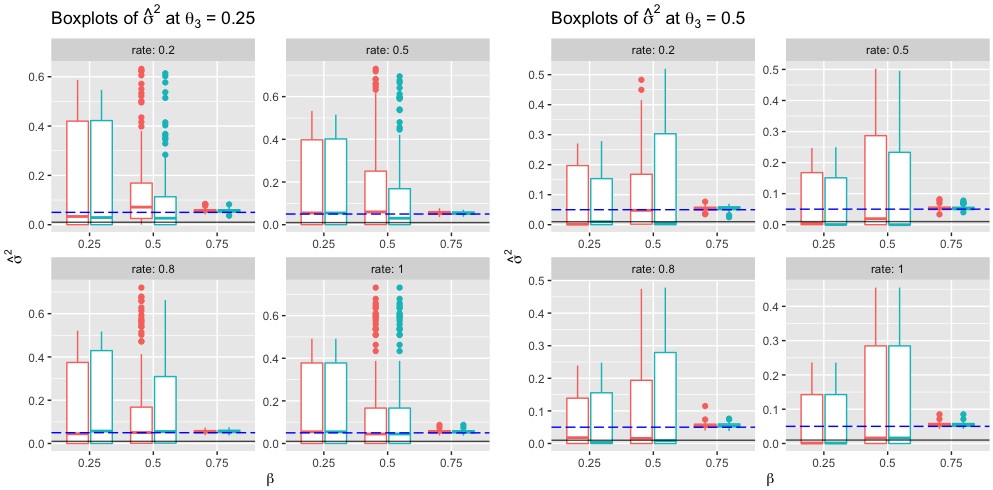


FIGURE 3.12. Boxplots of $\hat{\sigma}^2$. The solid reference line is the true parameter (0.01), and the dashed line is an initial point for the optimization. Red is from random observed locations, green is from block locations

unstructured noise. We observe that the unstructured variance, σ^2, is difficult to be estimated adequately. We consider possible intrinsic issues for having accurate estimates of θ_3 and σ^2 simultaneously. For example, the true σ^2 is relatively small compared to the range of eigenvalues of the graph Laplacian or other parameters. It would be more challenging to estimate the σ^2 in a less smooth process, having a smaller θ_3.

We generate 100 Gaussian processes on an identical graph with four parameter sets,

- $\Theta_1 = (\theta_1 = 1, \theta_2 = 1, \theta_3 = 0.25)$
- $\Theta_2 = (\theta_1 = 0.1, \theta_2 = 1, \theta_3 = 0.25)$
- $\Theta_3 = (\theta_1 = 1, \theta_2 = 1, \theta_3 = 0.75)$
- $\Theta_4 = (\theta_1 = 0.1, \theta_2 = 1, \theta_3 = 0.75)$

and the two different variances of random noises, σ^2, are 0.01 and 0.1. We have eight different Gaussian processes on the identical graph (DPP, $N = 1012$, $\beta = 0.5$). We conduct grid searches for finding MLEs with the given θ_1 and κ. We investigate that the estimate of σ^2 gets closed to the true parameters at the larger θ_3, the large magnitude of unstructured fluctuations would be easier to be detected, and how the value of range parameters θ_1 might affect on the estimates of θ_3 and σ^2.

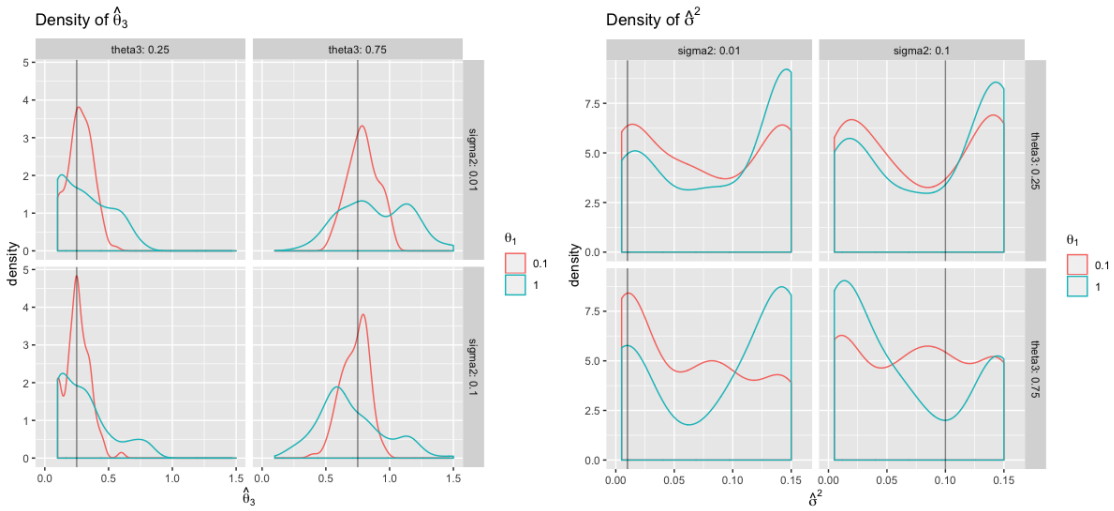


FIGURE 3.13. The solid line indicates the true parameters. The MLE of θ_3 are accumulated near the truth, but σ^2 is spread across the range of grid.

The magnitude of the range parameter, θ_1, affects the estimates of θ_3, see Figure 3.13. We confirm that the smaller θ_3 is easier to be estimated than larger θ_3 overall. Regardless of our settings, σ^2 is still difficult to estimate for both cases, $\sigma^2 = 0.1$ and 0.01.

We further examine the eigenvalues of $\Sigma(\boldsymbol{\theta})$, the covariance of the Gaussian process to understand the above behavior of estimates. The range of eigenvalues is from 0.5 to 1, and its approximate square root is 0.7. The true σ^2 is much smaller than the square root of eigenvalues, so the generated unstructured noise with σ^2 is less effective for the primary process due to the small magnitude. As we look at the eigenvalues of $\Sigma(\boldsymbol{\theta})$, we can see that θ_1 influences the shape of lines. The smaller θ_1 gives a smoother process and more informative eigenvalues.

Given (θ_2, θ_3) and with θ_1 treated as known, the σ^2 parameters are easier to estimate with the situation when θ_1 is smaller. Let $\theta_1' < \theta_1''$, θ_1' is easier to estimate (θ_2, θ_3) because there is greater slope with eigenvalue profile of $\Sigma(\boldsymbol{\theta})$.

MOM Approach for the unstructured variance. We consider the method of moment (MOM) estimator for σ^2 and illustrate the variance's behaviors upon the parameter sets.

The theoretical MOM estimator for σ^2 is as follows:

$$\hat{\sigma}^2_{MOM} = \frac{1}{N}\Big\{\text{tr}(\hat{\boldsymbol{\Gamma}}) - \text{tr}(\boldsymbol{\Gamma}(\boldsymbol{\theta}, \sigma^2))\Big\}, \tag{3.43}$$

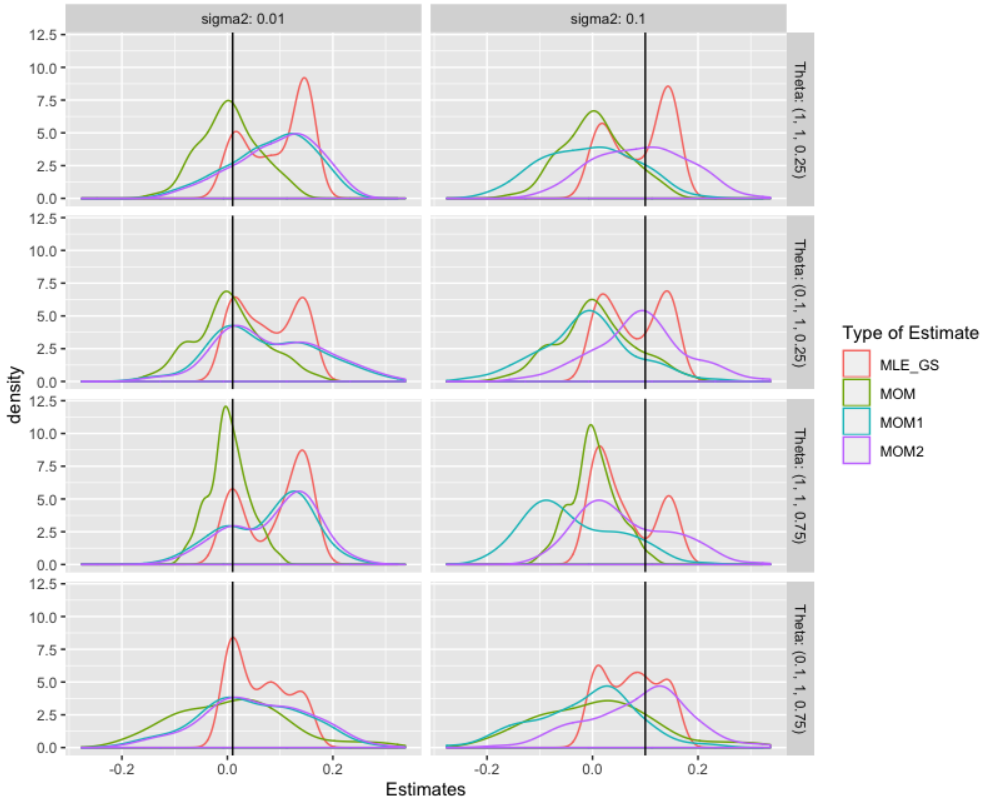


FIGURE 3.14. MLE_GS is an MLE for σ^2 from the previous grid search. The grid is constrained to greater than 0. MLE_GS concentrates on the boundaries of the grid.

where $\hat{\mathbf{\Gamma}}$ is a sample covariance from the full observations. MOM (3.43) is calculable when all true parameters are known. We define two other MOM scheme estimators for the unknown parameters.

$$\hat{\sigma}^2_{MOM,1} = \frac{1}{N}\Big\{\text{tr}(\hat{\mathbf{\Gamma}}) - \text{tr}(\mathbf{\Gamma}(\hat{\boldsymbol{\theta}}, \sigma^2))\Big\}, \tag{3.44}$$

$$\hat{\sigma}^2_{MOM,2} = \frac{1}{N}\Big\{\text{tr}(\hat{\mathbf{\Gamma}}) - \text{tr}(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}))\Big\}. \tag{3.45}$$

MOM1 is calculated with the known σ^2 and the estimated θ. MOM2 is evaluated with only the estimated $\boldsymbol{\theta}$ and assuming all parameters are unknown. The estimated θ are MLEs from the grid search results.

At a smaller $\sigma^2 = 0.01$, MOMs are centered around the true values. MOM1 and MOM2 tend to be centered at $\theta_3 = 0.25$. The qualities of MOM1 and MOM2 are dependent on the estimated

MLEs and the parameter $\boldsymbol{\theta}$. At a larger $\sigma^2 = 0.1$, MOM2 tends to be superior to the others. MOM2 occasionally produces an MLE closer to the truth.

3.5. Applications

We apply the proposed model to Gross Domestic Product per capita (GDPc) for a collection of its neighboring counties. We consider GDPc as a spatial process on the neighborhood county graph. We construct the graph by defining counties as nodes and the adjacency as edges.

We accessed U.S. county-level GDP data from the U.S. Bureau of Economic Analysis (BEA) and calculated GDPc by using the county population data from the U.S. Census. We obtained adjacency county information from the National Bureau of Economic Research (NBER).

We model the GDPc data from 2010 to 2018 in the Mid-Atlantic region (MA). MA includes New York, New Jersey, Pennsylvania, Delaware, Maryland, West Virginia, Virginia, and Washington DC, and it has 314 counties and independent cities, which are equivalent to counties in Virginia. We take partial information from 80% of the nodes to estimate the model and evaluate predictions.

3.5.1. County Graph. We create an undirected equal-weighted graph from the adjacency of counties in the MA region. We treat each county as a node, and the node is connected to the neighborhood county with equal weight. The MA graph has 314 nodes and 824 edges, and it is generated based on the available population and GDPc data from 2010 to 2018.

Our covariance structure is based on the eigenvalues and eigenvectors of the graph Laplacian. Specific dominant values sensitively influence the eigenvalues, so we have removed seven counties from the graph, which have extreme GDPc or variability over time. The new graph has 307 nodes and 792 edges; see Figure 3.15. Its degree distribution is symmetric, and the eigenvalues are smoothly decreasing; see Figure 3.16. The few dominant eigenvalues indicate that few nodes are connected to many nodes. There are two zero eigenvalues due to the removal of extreme values, so we have two unconnected nodes. We use the new graph for our GDPc modeling.

3.5.2. GDPc. Let $Y_t(\nu)$ be GDPc observation at county ν and time t. The model is as follows:

$$Y_t(\nu) = \mu_t + \mu_t(\nu) + X_t(\nu) + \epsilon_t(\nu), \qquad \nu \in \mathcal{G}, t = 1, \cdots, T, \tag{3.46}$$

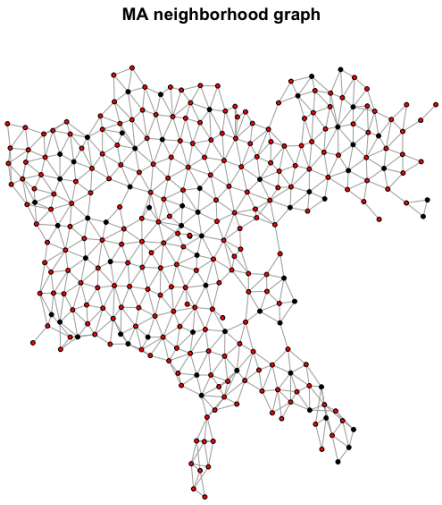


FIGURE 3.15. The neighborhood graph of MA region. We randomly select 80% observations (Red) to estimate the model.

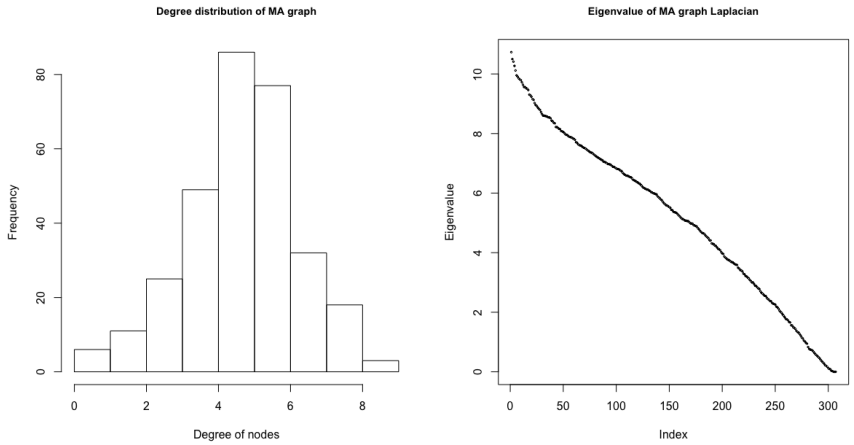


FIGURE 3.16. Degree distribution of MA graph and Eigenvalues of Laplacian of MA graph

where μ_t is a yearly mean for each t, $\mu_t(\nu)$ is a county-wise mean, and $X_t(\nu)$ is the graph process as defined in (3.1). The unstructured noise $\epsilon_t(\nu)$ are independent and identically distributed Gaussian with mean zero and variance σ^2.

We denote the residual fluctuation as $y_t(\nu)$.

$$y_t(\nu) \sim N(0, \mathbf{\Gamma}(\boldsymbol{\theta}, \sigma^2), \tag{3.47}$$

where $\mathbf{\Gamma}(\boldsymbol{\theta}, \sigma^2) = \mathbf{\Sigma}(\boldsymbol{\theta}) + \sigma^2 I$, $\mathbf{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\chi} g(\boldsymbol{\theta}, \Lambda) \boldsymbol{\chi}^T$. $g(\boldsymbol{\theta}, \Lambda)$ is the graphical Matérn model (3.6). Also Λ and $\boldsymbol{\chi}$ are eigenvalues and eigenvectors of the MA graph Laplacian.

We consider a Laplace (LP) transform of $y_t(\nu)$ as $\text{sign}(y_t(\nu))\sqrt{|y_t(\nu)|}$ as well. The distribution of our observed $y_t(\nu)$ is a sharper peak and thin longer tail; see Figure 3.17. We conduct modeling on the $y_t(\nu)$ and the LP transform.

Test Stationarity. We conduct the two tests in Subsection 3.2.3 to check the stationary model assumption of our observations. We calculate sample covariance matrices from observations and then use the transformed covariance matrix (in the eigenbasis of the graph Laplacian) to compute the test statistics. By the complete independence test, we set the null hypothesis as no correlation across observations for each node. We reject the null at $\alpha = 0.01$. We test the covariance structure with the null hypothesis as there is no correlation if the nodes are not directly connected. We fail to reject the null at $\alpha = 0.01$. From the tests, one may conclude that our observations are not independent but stationary regarding their nodes.

3.5.3. Estimation and Prediction. We estimate θ_2, θ_3 and σ^2 with a given θ_1 and a fixed $\kappa = 0.5$ in in-sample. The in-sample has observations on 245 nodes (80% of 307 nodes), which are randomly chosen, and the rest are assigned to out-of-sample. We implement Nelder-Mead optimization in `R` for searching θ_2, θ_3 and σ^2 simultaneously with a given θ_1 of $(0.1, 0.5, 1)$

We compute mean squared error (MSE) and predicted squared error (PMSE), which are the quantity $((Y - \widetilde{Y})^2/N)$, respectively, for in-sample and out-of-sample prediction problems, where $\widetilde{Y}$ denotes the best linear unbiased predictor (BLUP) of Y based on the holdout observations.

We also evaluate the quality of the model fit by considering the Frobenius norm of the normalized difference between the fitted covariance and the sample covariance matrix, namely, $\|\hat{\mathbf{\Gamma}}^{-1} S - I\|_F$, where S is corresponding the sample covariance matrix. We report the results in Table 3.1.

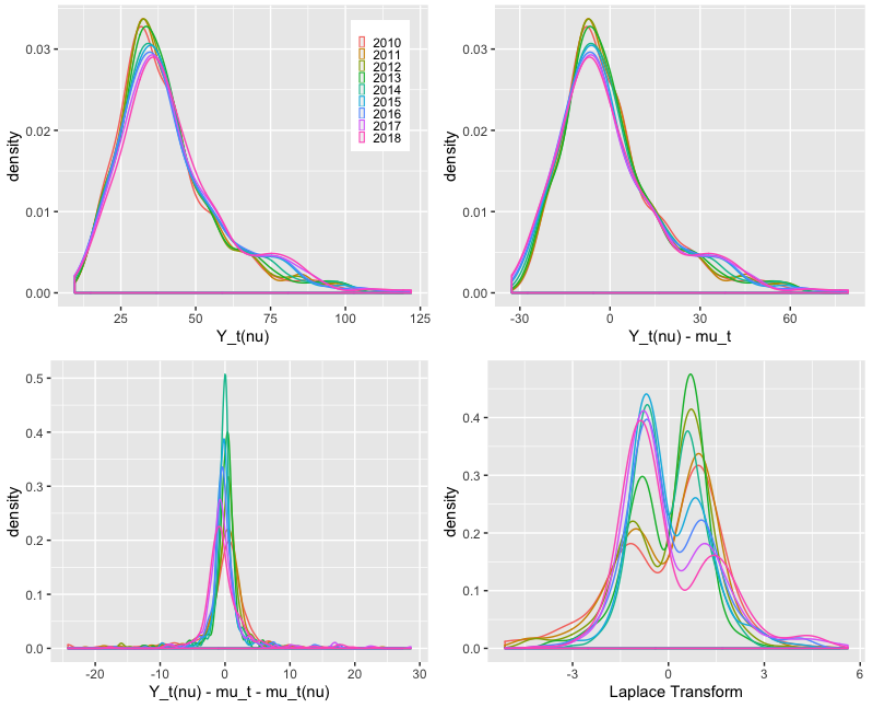


FIGURE 3.17. The GDPc data. The bottom illustrates the data set after removing yearly mean μ_t and after removing county-wise mean $\mu_t(\nu)$.

	θ_1	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\sigma}^2$	NLL	MSE	PMSE	Frob. Diff.	LP
1	1.0	1.204	2.103	1.168	1169.563	10.686	12.871	349.379	Y
2	0.5	0.524	1.264	1.130	1169.822	10.496	12.876	349.710	Y
3	0.5	0.553	1.220	1.117	1169.831	10.384	12.870	349.090	Y
4	0.1	0.563	0.438	0.905	1170.385	8.737	12.888	350.323	Y
5	0.1	1.043	0.247	0.457	1170.549	3.411	12.874	349.457	Y
6	1.0	4.085	0.687	3.133	2366.071	4.048	12.001	657.048	N
7	1.0	4.328	0.596	2.821	2366.078	3.280	11.998	657.025	N
8	1.0	6.810	0.272	0.073	2366.196	0.002	12.009	657.337	N
9	0.5	5.778	0.227	0.383	2366.286	0.061	12.028	658.165	N
10	0.5	6.146	0.211	0.013	2366.286	0.000	12.028	658.167	N
11	0.5	5.207	0.257	0.956	2366.287	0.378	12.028	658.171	N
12	0.5	2.509	0.651	3.619	2366.546	5.413	12.085	658.648	N
13	0.1	5.558	0.140	0.006	2367.051	0.000	12.152	662.808	N
14	0.1	5.529	0.143	0.042	2367.054	0.001	12.145	662.444	N
15	0.1	4.712	0.166	0.842	2367.114	0.297	12.158	662.912	N
16	0.1	4.397	0.177	1.148	2367.145	0.552	12.164	663.466	N
17	1.0	0.127	0.210	4.963	2372.258	11.167	12.970	694.511	N
18	0.1	0.161	0.077	4.914	2372.272	10.932	12.972	694.006	N

TABLE 3.1. The result table is arranged by negative log likelihood (NLL). LP refers the Laplace transform.

We select the 7th model in the Table 3.1 as the final model by the minimum PMSE criterion. Because of the scale difference between the LP transform data and the original $y_t(\nu)$, LP transforms have smaller negative log-likelihood (NLL) and the Frobenius difference. NLL and Frobenius difference can only be used within an identical transformation. PMSE allows us to compare all models across transformations. The estimate of σ^2 is closed to its MSE, which indicates the model is closed to a true model.

We gain some understanding of the GPDc process in the neighborhood graph with our final model. We gain confirmation that the scale of the processes is related to connectivity to other countries because the estimate of the scale parameter is in the middle range of our eigenvalues of the MA graph Laplacian. The smoothness parameter indicates that the correlations are not rapidly decreasing across the nodes.

3.6. Discussion

We propose a stationary Gaussian process on undirected graphs. It provides a new tool for the inference of processes on an undirected graph using spectral graph theory. Processes observed on a set of nodes generalize stationary processes from temporal or spatial domains to the setting of the undirected graph.

We characterize the processes using the graph Laplacian and study how the graph structure connects to modeling the process on the graph. In simulation studies, the estimates are more clustered around the true parameters as the size of the node increases, which indicates the consistency of our estimates. We find that the node configuration and connectivity contributes significantly to our model because they are associated with the eigenvalues and eigenvectors. The subgraph approach produces more accurate estimation and prediction for the observations. In the graphical Matérn model, it is challenging to obtain MLEs of the range parameter and the unstructured variance and find feasible initial values for the optimization method. In application, we model the GDPc on a county graph. Our model provides a framework in the context of socio-economic data.

We have done some preliminary research by utilizing the characteristics of the eigenvalues of the graph Laplacian for the geometric graph. We plan to extend our research to use the weighted graph and other types of parametric spectral density functions. In application, we may consider

different ways to construct a graph and embed it with processes by "diffusion map" approaches [31], [37]. We further investigate the roles of eigenvectors regarding observed processes.

Bibliography

[1] Apanasovich, T. V., Genton, M. G. and Sun, Y. [2012], 'A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components', *Journal of the American Statistical Association* **107**(497), 180–193.

[2] Barrera, R. G., Estevez, G. A. and Giraldo, J. [1985], 'Vector spherical harmonics and their application to magnetostatics', *European Journal of Physics* **6**(4), 287–294.

[3] Belkin, M. and Niyogi, P. [2008], 'Towards a theoretical foundation for Laplacian-based manifold methods', *Journal of Computer and System Sciences* **74**, 1289–1308.

[4] Breiman, L. [2001], 'Random forests', *Machine Learning* **45**(1), 5–32.

[5] Cai, T. T. and Jiang, T. [2011], 'Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices', *Ann. Statist.* **39**(3), 1496–1525.

[6] Chulliat, A. S. M. P. A. C. B. M. N. B. H. A. W. V. R. S. M. A. T. [2015], The us/uk world magnetic model for 2015-2020, Technical report, National Geophysical Data Center, NOAA.

[7] Chung, F. R. K. [1997], *Spectral Graph Theory*, CBMS Reigonal Conference Series in Mathematics, American Mathematical Society, Providence, RI.

[8] Driscoll, J. R. and Healy, D. M. [1994], 'Computing Fourier transforms and convolutions on the 2-sphere', *Advances in Applied Mathematics* **15**(2), 202–250.

[9] Efron, B. [1987], 'Better bootstrap confidence intervals', *Journal of the American Statistical Association* **82**(397), 171–185.

[10] Fan, M., Paul, D., Lee, T. C. M. and Matsuo, T. [2018], 'Modeling tangential vector fields on the sphere', *Journal of the American Statistical Association* **113**, 1625–1636.

[11] Finlay, C. C., Olsen, N., Kotsiaros, S., Gillet, N. and Tøffner-Clausen, L. [2016], 'Recent geomagnetic secular variation from swarm and ground observatories as estimated in the chaos-6 geomagnetic field model', *Earth, Planets and Space* **68**(1), 112.

[12] Finlay, C. C., Olsen, N. and Tøffner-Clausen, L. [2015], 'DTU candidate field models for IGRF-12 and the CHAOS-5 geomagnetic field model', *Earth, Planets and Space* **67**(114).

[13] Freeden, W. and Schreiner, M. [2009], *Spherical Functions Of Mathematical Geosciences : A Scalar, Vectorial, And Tensorial Setup*, Advances in Geophysical and Environmental Mechanics and Mathematics, Springer-Verlag.

[14] Fuselier, E. and Wright, G. [2009], 'Stability and error estimates for vector field interpolation and decomposition on the sphere with rbfs', *SIAM Journal on Numerical Analysis* **47**(5), 3213–3239.

[15] Gneiting, T., Kleiber, W. and Schlather, M. [2010], 'Matérn cross-covariance functions for multivariate random fields', *Journal of the American Statistical Association* **105**(491), 1167–1177.

[16] Hall, P., Kay, J. W. and Titterington, D. M. [1991], 'On estimation of noise variance in two-dimensional signal processing', *Advances in Applied Probability* **23**(3), 476–495.

[17] Healy Jr., D. M., Rockmore, D. N., Kostelec., P. J. and Moore, S. [2003], 'FFTs for the 2-sphere–improvements and variations', *Journal of Fourier Analysis and Applications* **9**(4), 341–385.

[18] Hill, E. L. [1954], 'The theory of vector spherical harmonics', *American Journal of Physics* **22**(4), 211–214.

[19] Hoffmeyer, P. [2000], 'The oersted satellite project', *Air and Space Europe* **2**(5), 74 – 79.

[20] Holme, R. [2000], 'Modelling of attitude error in vector magnetic data: application to ørsted data', *Earth, Planets and Space* **52**(12), 1187–1197.

[21] Holme, R. and Bloxham, J. [1996], 'The treatment of attitude errors in satellite geomagnetic data', *Physics of the Earth and Planetary Interiors* **98**(3), 221 – 233.

[22] Holton, J. R. and Hakim, G. J. [2013*a*], Chapter 3 - elementary applications of the basic equations, *in* J. R. Holton and G. J. Hakim, eds, 'An Introduction to Dynamic Meteorology (Fifth Edition)', fifth edition edn, Academic Press, Boston, pp. 67 – 93.

[23] Holton, J. R. and Hakim, G. J. [2013*b*], Chapter 4 - circulation, vorticity, and potential vorticity, *in* J. R. Holton and G. J. Hakim, eds, 'An Introduction to Dynamic Meteorology (Fifth Edition)', fifth edition edn, Academic Press, Boston, pp. 95 – 125.

[24] Hulot, G., Sabaka, T., Olsen, N. and Fournier, A. [2015], 5.02 - the present and future geomagnetic field, *in* G. Schubert, ed., 'Treatise on Geophysics (Second Edition)', second edition edn, Elsevier, Oxford, pp. 33 – 78.

[25] Jun, M. [2011], 'Non-stationary cross-covariance models for multivariate processes on a globe', *Scandinavian Journal of Statistics* **38**(4), 726–747.

[26] Jun, M. [2014], 'Matérn-based nonstationary cross-covariance models for global processes', *Journal of Multivariate Analysis* **128**, 134–146.

[27] Jun, M. and Stein, M. L. [2008], 'Nonstationary covariance models for global data', *The Annals of Applied Statistics* **2**(4), 1271–1289.

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

[29] Kostelec, P. J., Maslen, D. K., Healy, D. M. and Rockmore, D. N. [2000], 'Computational harmonic analysis for tensor fields on the two-sphere', *Journal of Computational Physics* **162**(2), 514 – 535.

[30] Kostelec, P. J. and Rockmore, D. N. [2004], S2kit: A lite version of spharmonickit.

[31] Lafon, S. [2004], Diffusion Maps and Geometric Harmonics, PhD thesis, Yale University.

[32] Langel, R., Sabaka, T., Baldwin, R. and Conrad, J. [1996], 'The near-earth magnetic field from magnetospheric and quiet-day ionospheric sources and how it is modeled', *Physics of the Earth and Planetary Interiors* **98**(3), 235 – 267.

[33] Lavancier, F., Møller, J. and Rubak, E. [2015], 'Determinantal point process models and statistical inference', *Journal of the Royal Statistical Society: Series B: Statistical Methodology* pp. 853–877.

[34] Marinucci, D. and Peccati, G. [2011], *Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications*, Vol. 389 of *London Mathematical Society Lecture Note Series*, Cambridge University Press.

[35] Maus, S., Luhr, H., Rother, M., Hemant, K., Balasis, G., Ritter, P. and Stolle, C. [2007], 'Fifth-generation lithospheric magnetic field model from CHAMP satellite measurements', *Geochemistry, Geophysics, Geosystems* **8**(5).

[36] Maus, S., Manoj, C., Rauberg, J., Michaelis, I. and Lühr, H. [2010], 'NOAA/NGDC candidate models for the 11th generation International Geomagnetic Reference Field and the concurrent release of the 6th generation Pomme magnetic model', *Earth, Planets and Space* **62**(10), 729–735.

[37] Nadler, B., Lafon, S., Coifman, R. R. and Kevrekidis, I. G. [2006], 'Diffusion maps, spectral clustering and reaction coordinates of dynamical systems', *Applied and Computational Harmonic Analysis* **21**, 113–127.

[38] Narcowich, F. J., Ward, J. D. and Wright, G. B. [2007], 'Divergence-free RBFs on surfaces', *Journal of Fourier Analysis and Applications* **13**, 643–663.

[39] Olsen, N., Holme, R., Hulot, G., Sabaka, T., Neubert, T., Tøffner-Clausen, L., Primdahl, F., Jørgensen, J., Léger, J.-M., Barraclough, D., Bloxham, J., Cain, J., Constable, C., Golovkov, V., Jackson, A., Kotzé, P., Langlais, B., Macmillan, S., Mandea, M., Merayo, J., Newitt, L., Purucker, M., Risbo, T., Stampe, M., Thomson, A. and Voorhies, C. [2000], 'Ørsted initial field model', *Geophysical Research Letters* **27**(22), 3607–3610.

[40] Olsen, N., Hulot, G. and Sabaka, T. J. [2010], *Sources of the Geomagnetic Field and the Modern Data That EnableTheir Investigation*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 105–124.

[41] Olsen, N., Lühr, H., Finlay, C. C., Sabaka, T. J., Michaelis, I., Rauberg, J. and Tøffner-Clausen, L. [2014], 'The CHAOS-4 geomagnetic field model', *Geophysical Journal International* **197**(2), 815–827.

[42] Ridgeway, G. [2005], Generalized boosted models : A guide to the gbm package.

[43] Sabaka, T. J., Hulot, G. and Olsen, N. [2010], Mathematical properties relevant to geomagnetic field modeling, *in* W. Freeden, M. Z. Nashed and T. Sonar, eds, 'Handbook of Geomathematics', Springer Berlin Heidelberg, pp. 503–538.

[44] Schott, J. R. [2005], 'Testing for complete independence in high dimensions', *Biometrika* **92**(4), 951–956.

[45] Singer, A. [2006], 'From graph to manifold Laplacian : the convergence rate', *Applied and Computational Harmonic Analysis* **21**, 128–134.

[46] Stein, M. L. [2007], 'Spatial variation of total column ozone on a global scale', *The Annals of Applied Statistics* **1**(1), 191–210.

[47] Watterson, I. G. [2001], 'Decomposition of global ocean currents using a simple iterative method', *Journal of Atmospheric and Oceanic Technology* **18**(4), 691–703.