Modeling for Spatial and Spatio-Temporal Data With Applications

by

Xintong Li

M.S., California State University, East Bay, 2013

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

Doctor of Philosophy

Department of Statistics
College of Arts and Sciences

Kansas State University
Manhattan, Kansas

2018

Abstract

It is common to assume the spatial or spatio-temporal data are realizations of underlying random fields or stochastic processes. Effective approaches to modeling of the underlying autocorrelation structure of the same random field and the association among multiple processes are of great demand in many areas including atmospheric sciences, meteorology and agriculture. To this end, this dissertation studies methods and application of the spatial modeling of large-scale dependence structure and spatio-temporal regression modelling.

First, variogram and variogram matrix functions play important roles in modeling dependence structure among processes at different locations in spatial statistics. With more and more data collected on a global scale in environmental science, geophysics, and related fields, we focus on the characterizations of the variogram models on spheres of all dimensions for both stationary and intrinsic stationary, univariate and multivariate random fields. Some efficient approaches are proposed to construct a variety of variograms including simple polynomial structures. In particular, the series representation and spherical behavior of intrinsic stationary random fields are explored in both theoretical and simulation study. The applications of the proposed model and related theoretical results are demonstrated using simulation and real data analysis.

Second, knowledge of the influential factors on the number of days suitable for fieldwork (DSFW) has important implications on timing of agricultural field operations, machinery decision, and risk management. To assess how some global climate phenomena

such as El Nino Southern Oscillation (ENSO) affects DSFW and capture their complex associations in space and time, we propose various spatio-temporal dynamic models under hierarchical Bayesian framework. The Integrated Nested Laplace Approximation (INLA) is used and adapted to reduce the computational burden experienced when a large number of geo-locations and time points is considered in the data set. A comparison study between dynamics models with INLA viewing spatial domain as discrete and continuous is conducted and their pros and cons are evaluated based on multiple criteria. Finally a model with time- varying coefficients is shown to reflect the dynamic nature of the impact and lagged effect of ENSO on DSFW in US with spatio-temporal correlations accounted.

Modeling for Spatial and Spatio-Temporal Data

With Applications

by

Xintong Li

M.S., California State University, East Bay, 2013

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

Doctor of Philosophy

Department of Statistics
College of Arts and Sciences

Kansas State University
Manhattan, Kansas

2018

Approved by:

Major Professor
Dr. Juan Du

Copyright

Abstract

It is common to assume the spatial or spatio-temporal data are realizations of underlying random fields or stochastic processes. Effective approaches to modeling of the underlying autocorrelation structure of the same random field and the association among multiple processes are of great demand in many areas including atmospheric sciences, meteorology and agriculture. To this end, this dissertation studies methods and application of the spatial modeling of large-scale dependence structure and spatio-temporal regression modelling.

First, variogram and variogram matrix functions play important roles in modeling dependence structure among processes at different locations in spatial statistics. With more and more data collected on a global scale in environmental science, geophysics, and related fields, we focus on the characterizations of the variogram models on spheres of all dimensions for both stationary and intrinsic stationary, univariate and multivariate random fields. Some efficient approaches are proposed to construct a variety of variograms including simple polynomial structures. In particular, the series representation and spherical behavior of intrinsic stationary random fields are explored in both theoretical and simulation study. The applications of the proposed model and related theoretical results are demonstrated using simulation and real data analysis.

Second, knowledge of the influential factors on the number of days suitable for fieldwork (DSFW) has important implications on timing of agricultural field operations, machinery decision, and risk management. To assess how some global climate phenomena

such as El Nino Southern Oscillation (ENSO) affects DSFW and capture their complex associations in space and time, we propose various spatio-temporal dynamic models under hierarchical Bayesian framework. The Integrated Nested Laplace Approximation (INLA) is used and adapted to reduce the computational burden experienced when a large number of geo-locations and time points is considered in the data set. A comparison study between dynamics models with INLA viewing spatial domain as discrete and continuous is conducted and their pros and cons are evaluated based on multiple criteria. Finally a model with time- varying coefficients is shown to reflect the dynamic nature of the impact and lagged effect of ENSO on DSFW in US with spatio-temporal correlations accounted.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to acknowledge and extend my greatest gratitude to my Major Professor Dr. Juan Du. She is a most knowledgable guide and a everlasting light of inspiration for all my research life. In two and half years, I have encountered many challenges in this research. It is her firm believe and help to keep me going and going in this route. Without her, this dissertation would not have been possible.

I also would like to thank my committee members, Dr. Griffin, Dr. Reyes and Dr. Song. Dr. Griffin provided us the original data set in Chapter 2. Through his guidance, I came to know the experiment design and data collection process of Day Suitable for Field Work. Even through Dr. Du guided me through this dissertation, it was Dr. Reyes first introduced me to the kingdom of Bayesian Hierarchical Model. Through her unique angle, I was able to glance into the most frontier method in spatial/spatio-temporal statistics. I also want to appreciate Dr. Song for his statistics in the High-dimension class. His class taught me how to read a research paper and improve my skills in theorem proof.

In addition, I would like to thank Dr. Hefley for his extensive knowledge in application of spatial/ spatio-temporal statistics; Dr. Jun for the insight of spatial model on spheres and Dr. Zidek for the introducing INLA method.

At last, many thanks to my family and friends who kept encouraging me through out my studies.

Dedication

This is dedicated to my grand parents, Yumin Li and Yihua Liu. For all the love and support they have given to me over twenty years; For believing me of being able to accomplish more than what I had; For making me smile, laugh, sad and cry when raising me from a six-years old child to a person big enough to take responsibilities; For waiting to talk to me every weekend in the past eight years; and For remembering me even after Alzheimer came to you and you almost forgot everything, you didn't forget me. I couldn't have done it without you. I have all your love and you will have all mine.

And to my Mom and Dad, Aunt and Uncle, and my cousins. My family makes me strong and I dedicate this work to them.

Preface

In modern science, more and more data is collected on a large scale. Not only the number of locations and time points becomes large, it also cover a great area. Model on spheres in spatial statistics is to capture spatial dependent structure over a globe, which may applied to many applications in geophysics and climatology, such as global mapping of greenhouse gas fluxes, temperature-precipitation interactions in climate change and so on. Since we are working on distance on spheres. The distortion between Euclidean distance and spherical distance can be not ignored. We must validate positive definite properties of the variogram constructed from this study. A series of theorems and corollaries are developed for this purpose. In this dissertation, we also focus on variogram and variogram matrix function on spheres of all dimensions. Some intrinsically stationary random fields may not have a valid covariance structure by definition, but variogram can be a relevant tool in modeling dependence structures among those fields. We also found many data collection process on spheres are multivariate by their nature. In particular, the characterizations of isotropic variogram models on all spheres for both univariate and multivariate random fields are engaged and some efficient approach are proposed with a simple polynomial structure. Furthermore, to compare our model with existing method fairly, we propose a new simulation method, which allow us to generate a multivariate intrinsic stationary, Gaussian or Non-Gaussian spatial random field.

In chapter 2, we focus on methods and applications of Bayesian dynamic models on spatio-temporal data. One application in agriculture economics is used as a source of our question to demonstrate how to solve a complex questions in dynamic models' framework.

Days suitable for fieldwork (DSFW) is the count of farming operation within a week or a month. The El Niño Southern Oscillation (ENSO) is a naturally occurring phenomenon that involves fluctuation of ocean temperatures in the equatorial Pacific. The knowledge of how global weather pattern such as ENSO influence the number of days suitable for fieldwork (DSFW) has important implications on timing of agricultural field operations, machinery decision, and risk management. The Bayesian dynamic models are used to assess the complex association between ENSO and DSFW in space and time. However, the computational burden of large amount of data in space and time has to be considered thoroughly. Integrated Nested Laplace Approximation perform fast full Bayesian inference through accurate approximation of marginal posterior and latent variables. INLA reduces computational cost from days in MCMC to hours and then allow us to fit some very complex dynamic models in a large data set. Besides, we also use DSFW and ENSO data as an example to investigate and compare dynamic models in discrete spatial domain and continuous spatial domain. A lagged effect model with time-varying coefficients are finalized to make inferences.

Chapter 1

Variogram Models on Spheres of All Dimensions

1.1 Introduction

With more and more data collected on a global scale in geophysics, climatology and related fields, efficiently capturing spatial dependence structure on spheres play an increasingly important role in understanding the underlying processes, such as global mapping of greenhouse gas fluxes, temperature-precipitation interactions in climate change, and so on (see e.g. Adviento-Borbe et al. 2007, Gaspari and Cohn 1999, Sain et al. 2011, Du et al. 2013, among others). Recently, it has caught great attention to develop and validate covariance models on spheres (see e.g. Jun and Stein 2007, Jun and Stein 2008, Cressie and Johannesson 2008, Du and Ma 2011, Ma 2012, Du et al. 2013, Ma 2015), although the idea of positive definite functions on spheres was introduced as early as 1940's (Bochner 1941, Schoenberg 1942).

Comparing with covariogram or covariance matrix function, variogram or variogram

matrix function can be more general and relevant tools in modeling dependence structure among multiple processes at different locations in spatial statistics (Matheron 1963, Journel and Huijbregts 1978, Du and Ma 2011). For example, some intrinsically stationary random fields may not have finite second order moments and therefore covariograms do not exist, but variogram models might still be used for capturing the co-reaction of the attributes at different locations. In fact an isotropic variogram on a sphere may not only be associated with a stationary random field on the sphere but also an intrinsically stationary random field on the sphere, such as a fractional Brownian motion on a sphere. Moreover, the estimation of parameters through the variogram is more stable in univariate case, and is recommended for estimation in multivariate case (Cressie et al. 2015). It is also worth mentioning that kriging or cokriging can be achieved through variogram or variogram matrix models (Cressie 1993). Jeong and Jun (2015) demonstrate that the distortion between Euclidean distance and spherical distance is not negligible when the spatial range of the data is large and the models originally defined on Euclidean space are often not physically justifiable on spheres. They further argue the need to construct variogram specifically built for spheres, even when Matérn or other variogram can be valid on all spheres through chordal distance to spherical distance mapping. Before adapting Euclidean covariance and variogram models to the sphere, one must first ensure their validity (Huang et al. 2011). While the Gaussian random field on the Euclidean space is well studied (Lang and Schwab 2015), the study on variogram modeling and properties on all spheres is limited in literature. Most of them are focusing on univariate isotropic covariance for stationary random fields without actual implementation to show the usefulness of the given models in real data analysis (e.g. Ma 2015). Since data on the globe often cover a very large spatial range, the stationary assumption can be violated easily and long distance prediction is more likely due to large scale land zone variation.

Thus, intrinsically stationary or nonstationary models on spheres are of great demand, and so are the models that capture negative spatial correlation and hole effects. To deal with nonstationarity on spheres, kernel convolution has been effectively used to model underlying random field (see e.g. Heaton et al. 2014, Li and Zhu 2016). But the associated numerical calculation sometimes can be challenging with no closed form for its corresponding covariance functions generally. In this work we focus on explicit variogram modeling for intrinsically stationary random fields with straightforward calculation. In particular, we engage in the characterizations of isotropic variogram models on spheres of all dimensions for both univariate and multivariate random fields. More efficient approaches are introduced to construct isotropic variogram models with various structures including simple polynomial forms with hole effect. Some of which will prove to be efficient in modeling and effective on long distance kriging through simulation study and data analysis. Moreover, a new simulation method will be proposed to generate intrinsic stationary, Gaussian or non-Gaussian, multivariate spatial data, given the fact that most of the existing spatial simulation methods are based on Gaussian and weakly stationary assumptions.

The rest of this chapter is organized as follows. Sections 2 and 3 establish sufficient and necessary conditions for validating univariate variogram or multivariate variogram matrix functions for random fields with isotropic increments on all spheres. Those sufficient and necessary conditions can further be used to find an adequate form of variogram matrix functions when needed, and to provide a guideline to develop new forms of variograms on all spheres. In Section 4, we introduce a new method to simulate a vector random field on a sphere, which may or may not be a Gaussian random field. The simulation study in this section and data analysis in next section are conducted to show how to apply the proposed models and demonstrate the advantages of some models developed in certain

scenarios, such as for long distance prediction with separation of oceans or mountains. Particularly one proposed polynomial type of variogram turns out to be computationally efficient, valid on all spheres and flexible enough to fit different large scale data sets. Proofs of Theorems 1-5 are given in the Appendix.

1.2 Univariate Isotropic Variogram Models on All Spheres

Geophysical and environmental data are often viewed as a realization of a random field $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{S}^d\}$, where $\mathbb{S}^d$ is a sphere or spherical shell of radius one and centered at the origin in $\mathbb{R}^{d+1}$, i.e., $\mathbb{S}^d = \{\|\mathbf{x}\| = 1, \mathbf{x} \in \mathbb{R}^{d+1}\}$, where $\|\mathbf{x}\|$ is the Euclidean norm of x. For two points $\mathbf{x}_1$ and $\mathbf{x}_2$ on $\mathbb{S}^d$, their spherical (angular or great-circle) distance is denoted by $\vartheta(\mathbf{x}_1, \mathbf{x}_2),\ 0 \le \vartheta(\mathbf{x}_1, \mathbf{x}_2) \le \pi$. It is related to the Euclidean (chordal) distance $\|\mathbf{x}_1 - \mathbf{x}_2\|$ and the inner product $\mathbf{x}_1{}'\mathbf{x}_2$ through the following formulas,

$$\|\mathbf{x}_1 - \mathbf{x}_2\|^2 = 2(1 - \mathbf{x}_1{}'\mathbf{x}_2) = 2\{1 - \cos(\vartheta(\mathbf{x}_1, \mathbf{x}_2))\},\ \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d. \tag{1.1}$$

A random field $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{S}^d\}$ is said to have second order increments if its increments have second-order moments. Moreover, $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{S}^d\}$ is called an intrinsically stationary random field with isotropic increments, if $E(Z(\mathbf{x}_1) - Z(\mathbf{x}_2))$ does not depend on $\mathbf{x}_1$ and $\mathbf{x}_2$, and its variogram

$$\gamma(\vartheta(\mathbf{x}_1, \mathbf{x}_2)) = \frac{1}{2}\operatorname{var}(Z(\mathbf{x}_1) - Z(\mathbf{x}_2)) = \frac{1}{2}E[Z(\mathbf{x}_1) - Z(\mathbf{x}_2) - E(Z(\mathbf{x}_1) - Z(\mathbf{x}_2))]^2,\ \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d, \tag{1.2}$$

depends only on the spherical distance $\vartheta(\mathbf{x}_1, \mathbf{x}_2)$ between $\mathbf{x}_1$ and $\mathbf{x}_2$ on $\mathbb{S}^d$. We say $\gamma(\vartheta)$

is an isotropic variogam on all spheres or $\mathbb{S}^\infty$, if $\gamma(\vartheta), \vartheta \in [0,\pi]$, is an isotropic variogram on all $\mathbb{S}^d (d = 1, 2, ...)$.

A nonnegative and continuous function $g(x)$ is absolutely monotone on $[a, b]$, if it has non-negative derivatives of all orders over (a, b), i.e.,

$$g^{(n)}(x) \geq 0,\ a < x < b,\ n = 1, 2, \ldots.$$

The following theorem characterizes a nonnegative and continuous function to be an isotropic variogram on $\mathbb{S}^\infty$.

Theorem 1. *For a nonnegative and continuous function $\gamma(\vartheta)$ on $[0,\pi]$, the following statements are equivalent:*

(i) $\gamma(\vartheta)$ is an isotropic variogram on $\mathbb{S}^\infty$;

(ii) $\gamma(\vartheta)$ is of the form

$$\gamma(\vartheta) = \sum_{n=1}^{\infty} b_n \left(1 - \cos^n \vartheta\right),\ \vartheta \in [0,\pi], \tag{1.3}$$

where $\{b_n, n = 1, 2, \ldots\}$ is a summable sequence of nonnegative numbers;

(iii) $\gamma(\vartheta)$ is of the form

$$\gamma(\vartheta) = g(1) - g(\cos\vartheta),\ \vartheta \in [0,\pi], \tag{1.4}$$

where the function $g(x)$ is continuous on $[-1, 1]$ and absolutely monotone on $[0, 1]$, and $g(x) - g(-x)$ is absolutely monotone on $[0, 1]$;

(iv) $\gamma_0 - \gamma\left(\frac{\pi}{2} - \arcsin x\right)$ and $\gamma\left(\frac{\pi}{2} + \arcsin x\right) - \gamma\left(\frac{\pi}{2} - \arcsin x\right)$, $x \in [-1, 1]$, are absolutely monotone on $[0, 1]$, and $\gamma(0) = 0$, where γ_0 is a constant with $\gamma_0 \geq \max\limits_{0\leq\vartheta\leq\pi} \gamma(\vartheta)$.

Theorem 1 (iii) and (iv) above provide handy tools to validate variograms on $\mathbb{S}^\infty$, given a wealth of ways to construct and verify absolute monotone functions in the literature (Widder 1941, Feller 1971). In addition, Theorem 1 (ii) results in the following two properties of an isotropic variogram, which can be used to show that certain function is not a variogram on $\mathbb{S}^\infty$.

Corollary 1. *If $\gamma(\vartheta)$, $\vartheta \in [0, \pi]$ is an isotropic variogram on $\mathbb{S}^\infty$, then it is nonnegative on $[0, \pi]$, increasing on $[0, \frac{\pi}{2}]$, and $\gamma(0) = 0$.*

Corollary 1 is directly derived from theorem 1, since each b_n is nonnegative and $\cos^n \vartheta$ lies between -1 and 1. In view of Corollary 1, $\gamma(\vartheta)$ has to be increasing on $\left[0, \frac{\pi}{2}\right]$, but it may not be so on the interval $\left[\frac{\pi}{2}, \pi\right]$, as is released from Corollary 2.

Corollary 2. *If $\gamma(\vartheta)$, $\vartheta \in [0, \pi]$ is an isotropic variogram on $\mathbb{S}^\infty$, then $\gamma\left(\frac{\pi}{2} + \vartheta\right) - \gamma\left(\frac{\pi}{2} - \vartheta\right)$ is nonnegative and increasing on $\left[0, \frac{\pi}{2}\right]$, and $\gamma\left(\frac{\pi}{2} - \vartheta\right) + \gamma\left(\frac{\pi}{2} + \vartheta\right)$ is nonnegative and decreasing on $\left[0, \frac{\pi}{2}\right]$.*

In fact, it follows from identity (1.3) that

$$\begin{aligned}
\gamma\left(\frac{\pi}{2} + \vartheta\right) - \gamma\left(\frac{\pi}{2} - \vartheta\right) &= \sum_{n=1}^{\infty} b_n \left\{1 - \cos^n\left(\frac{\pi}{2} + \vartheta\right)\right\} - \sum_{n=1}^{\infty} b_n \left\{1 - \cos^n\left(\frac{\pi}{2} - \vartheta\right)\right\} \\
&= \sum_{n=1}^{\infty} b_n \left(-(-1)^n \sin^n \vartheta + \sin^n \vartheta\right) \\
&= 2\sum_{n=0}^{\infty} b_{2n+1} \sin^{2n+1} \vartheta, \ \vartheta \in [0, \frac{\pi}{2}],
\end{aligned}$$

which takes nonnegative values and increases on $\left[0, \frac{\pi}{2}\right]$, since $\sin\vartheta$ does so. Similarly,

$$\begin{aligned}
\gamma\left(\frac{\pi}{2}+\vartheta\right)+\gamma\left(\frac{\pi}{2}-\vartheta\right) &= \sum_{n=1}^{\infty} b_n\left\{1-\cos^n\left(\frac{\pi}{2}+\vartheta\right)\right\}+\sum_{n=1}^{\infty} b_n\left\{1-\cos^n\left(\frac{\pi}{2}-\vartheta\right)\right\} \\
&= \sum_{n=1}^{\infty} b_n\left(2-(-1)^n\sin^n\vartheta-\sin^n\vartheta\right) \\
&= 2\sum_{n=1}^{\infty} b_{2n}\left(1-\sin^{2n}\vartheta\right),\ \vartheta\in\left[0,\frac{\pi}{2}\right],
\end{aligned}$$

takes nonnegative value and decreases on $\left[0, \frac{\pi}{2}\right]$.

For the convenience of applications, we also give some sufficient conditions for variograms on all spheres in the sequel. We say a nonnegative and continuous function $g(x)$ is completely monotone on $[0, a)$, if it possesses derivatives of all order on $(0, a)$, and

$$(-1)^{(n)}g^{(n)}(x)\geq 0,\ 0<x<a,\ n=1,2,\ldots,$$

where a is a finite number or infinity. Obviously, a function that is completely monotone on $[0, a_1)$ is also completely monotone on a sub-interval $[0, a_2)$ with $0 < a_2 < a_1$, but it may not be so on an extended interval $[0, a_3)$ for $a_3 > a_1$. Theorem 2 (ii) below follows from Theorem 2 (i), and the latter can be derived from Theorem 1 (iv).

Theorem 2. *For a continuous and nonnegative function $\gamma(\vartheta)$ on $[0, \pi]$, with $\gamma(0) = 0$, it is an isotropic variogram on $\mathbb{S}^\infty$, if one of the following conditions is satisfied:*

(i) both $\gamma_0 - \gamma\left(\frac{\pi}{2} - x\right)$ and $\gamma\left(\frac{\pi}{2} + x\right) - \gamma\left(\frac{\pi}{2} - x\right)$ are absolutely monotone on $\left[0, \frac{\pi}{2}\right]$,

(ii) $\gamma_0 - \gamma(\vartheta)$ is completely monotone on $[0, \pi]$, where γ_0 is a constant such that $\gamma_0 \geq \max_{0\leq\vartheta\leq\pi}\gamma(\vartheta)$.

Theorem 2 along with preceding corollaries can be used to verify if a function on $[0, \pi]$ is a valid isotropic variogram on all spheres as shown in the following examples.

Example 1. *A simple linear power function*

$$\gamma(\vartheta) = b_0 + b_1\vartheta^{\lambda},\ \vartheta \in [0,\pi], \tag{1.5}$$

is a valid isotropic variogram on $\mathbb{S}^{\infty}$ *if and only if* $b_0 = 0$, $b_1 > 0$ *and* $0 < \lambda \leq 1$. *As a matter of fact, we obtain* $b_0 = 0$ *and* $b_1 \geq 0$ *directly from Corollary 1. In case* $b_1 > 0$, λ *must lie between 0 and 1. This follows from Corollary 2, since* $\gamma\left(\frac{\pi}{2} - \vartheta\right) + \gamma\left(\frac{\pi}{2} + \vartheta\right)$ *is decreasing on* $\left[0, \frac{\pi}{2}\right]$, *so that* $\gamma\left(\frac{\pi}{2} - 0\right) + \gamma\left(\frac{\pi}{2} + 0\right) \geq \gamma\left(\frac{\pi}{2} - \frac{\pi}{2}\right) + \gamma\left(\frac{\pi}{2} + \frac{\pi}{2}\right)$, *or* $2\left(\frac{\pi}{2}\right)^{\lambda} \geq \pi^{\lambda}$, *which implies* $\lambda \leq 1$. *On the other hand, when* $0 < \lambda \leq 1$, *(1.5) is a variogram on* $\mathbb{S}^{\infty}$, *since Theorem 2 (ii) applies with*

$$(-1)^n \frac{d(\gamma_0 - \gamma(\vartheta))}{d\vartheta^n} = (-1)^{(n+1)} \prod_{i=1}^{n} (\lambda - i + 1) b_1 \vartheta^{(\lambda - n)} \geq 0,\ n = 1, 2, \ldots.$$

Example 2. *For a non-zero constant* b_0, *a rational function*

$$\gamma(\vartheta) = \frac{b_0\vartheta}{1 + b_1\vartheta}, \vartheta \in [0,\pi], \tag{1.6}$$

is an isotropic variogram on $\mathbb{S}^{\infty}$ *if and only if* $b_0 > 0$ *and* $b_1 \geq 0$. *In fact, it follows from Corollary 1 that* $b_0 > 0$ *for* $\gamma(\vartheta)$ *to be increasing on* $\left[0, \frac{\pi}{2}\right]$ *and that* $b_1 > -\frac{1}{\pi}$ *for* $\gamma(\pi) \geq 0$. *If* $b_0 > 0$ *and* $0 > b_1 > -\frac{1}{\pi}$, *then*

$$\frac{d}{d\vartheta}\left(\gamma\left(\frac{\pi}{2} - \vartheta\right) + \gamma\left(\frac{\pi}{2} + \vartheta\right)\right) = -b_0\left\{1 + b_1\left(\frac{\pi}{2} - \vartheta\right)\right\}^{-2} + b_0\left\{1 + b_1\left(\frac{\pi}{2} + \vartheta\right)\right\}^{-2} > 0,$$

which makes Corollary 2 fail. If both b_0 *and* b_1 *are nonnegative, we can apply Theorem 2 (ii) to show that (1.6) is an isotropic variogram on* $\mathbb{S}^{\infty}$. *In this case,* $\gamma(\vartheta)$ *is increasing on* $[0,\pi]$ *and reaches its maximum at* $\vartheta = \pi$. *For* $\gamma_0 \geq \gamma(\pi)$, $\gamma_0 - \gamma(\vartheta)$ *is nonnegative and*

continuous on $[0,\pi]$*, and completely monotone, with*

$$\begin{aligned}(-1)^n\frac{d^n}{d\vartheta^n}(\gamma_0-\gamma(\vartheta)) &= (-1)^n\frac{d^n}{d\vartheta^n}\left(\gamma_0-\frac{b_0}{b_1}+\frac{b_0}{b_1(1+b_1\vartheta)}\right)\\ &= (-1)^n\frac{b_0}{b_1}\frac{d^n}{d\vartheta^n}\frac{1}{1+b_1\vartheta}\\ &= \frac{n!b_0b_1^{n-1}}{(1+b_1\vartheta)^{n+1}}\geq 0,\ \vartheta\in(0,\pi),\ n=1,2,\ldots,\end{aligned}$$

so that Theorem 2 (ii) applies.

Example 3. *For a positive* α*, a power exponential function*

$$\gamma(\vartheta)=1-\exp(-\alpha\vartheta^\nu),\ \vartheta\in[0,\pi], \tag{1.7}$$

is an isotropic variogram on $\mathbb{S}^\infty$ *if and only if* $\nu\in(0,1]$ *and* $\alpha>0$*. As is verified by Du and Ma (2012),* $\gamma(\vartheta)=1-\exp(-\alpha\vartheta^\nu),\vartheta\in[0,\pi]$*, is an istropic variogram on* $\mathbb{S}^\infty$*, under the conditions* $\nu\in(0,1]$ *and* $\alpha>0$*. For the necessary condition, by Corollary 1, the variogram being an increasing function on* $\left[0,\frac{\pi}{2}\right]$ *implies* $\alpha>0$*. By Corollary 2,*

$$\begin{aligned}&\frac{d}{d\vartheta}\left(\gamma(\frac{\pi}{2}-\vartheta)+\gamma(\frac{\pi}{2}+\vartheta)\right)\\ =\ &\alpha\nu\left(\frac{\pi}{2}+\vartheta\right)^{\nu-1}\exp\left(-\alpha\left(\frac{\pi}{2}+\vartheta\right)^\nu\right)-\alpha\nu\left(\frac{\pi}{2}-\vartheta\right)^{\nu-1}\exp\left(-\alpha\left(\frac{\pi}{2}-\vartheta\right)^\nu\right)\\ \leq\ &0,\ \vartheta\in\left[0,\frac{\pi}{2}\right],\end{aligned}$$

which implies,

$$\left(\frac{\pi}{2}+\vartheta\right)^{\nu-1}\exp\left(-\alpha\left(\frac{\pi}{2}+\vartheta\right)^\nu\right)\ \leq\ \left(\frac{\pi}{2}-\vartheta\right)^{\nu-1}\exp\left(-\alpha\left(\frac{\pi}{2}-\vartheta\right)^\nu\right),\ \vartheta\in\left[0,\frac{\pi}{2}\right].$$

To guarantee this inequality on $[0,\frac{\pi}{2}]$*, it is necessary that* $\nu\in(0,1]$*.*

Example 4. *For nonnegative constants λ, b_0 and b_1,*

$$\gamma(\vartheta) = b_0^{-\lambda} - (b_0 + b_1\vartheta)^{-\lambda},\ \vartheta \in [0, \pi], \tag{1.8}$$

is a variogram function on $\mathbb{S}^\infty$. This can be confirmed by applying Theorem 2 (ii), with

$$\begin{aligned}(-1)^n \frac{d^n}{d\vartheta^n}(\gamma_0 - \gamma(\vartheta)) &= (-1)^n \frac{d^n}{d\vartheta^n}(b_0 + b_1\vartheta)^{-\lambda} \\ &= n! b_1^n (b_0 + b_1\vartheta)^{-\lambda - n} \geq 0,\ \vartheta \in (0, \pi),\ n = 1, 2, \ldots,\end{aligned}$$

where γ_0 is a constant not less than $\gamma(\pi)$.

Example 5.

$$\gamma(\vartheta) = 1 - exp(-\alpha\vartheta^v), \vartheta \in [0, \pi], \tag{1.9}$$

is a valid variogram on $\mathbb{S}^\infty$ if and only if $v \in (0, 1]$ and $\alpha > 0$. As is verified by Du and Ma (2012), the stable variogram $\gamma(\vartheta) = 1 - exp(-\alpha\vartheta^v), \vartheta \in [0, \pi], v \in (0, 1], and\ \alpha > 0$ is a valid variogram on $\mathbb{S}^\infty$. Therefore, the sufficient condition is complete. For the necessary condition, by Corollary 1, the variogram is an increasing function implies $\alpha > 0$. By Corollary 2,

$$\begin{aligned}\frac{d(\gamma(\frac{\pi}{2} - \vartheta) + \gamma(\frac{\pi}{2} + \vartheta))}{d\vartheta} &= \alpha v(\frac{\pi}{2} + \vartheta)^{v-1} exp(-\alpha(\frac{\pi}{2} + \vartheta)^v) - \alpha v(\frac{\pi}{2} - \vartheta)^{v-1} exp(-\alpha(\frac{\pi}{2} - \vartheta)^v) \\ &\leq 0,\ for\ \vartheta \in [0, \frac{\pi}{2}]\end{aligned}$$

since α, v are positive numbers, which implies,

$$(\frac{\pi}{2} + \vartheta)^{v-1} exp(-\alpha(\frac{\pi}{2} + \vartheta)^v) \leq (\frac{\pi}{2} - \vartheta)^{v-1} exp(-\alpha(\frac{\pi}{2} - \vartheta)^v),\ for\ \vartheta \in [0, \frac{\pi}{2}]$$

To guarantee the inequality on the support of $\vartheta \in [0, \frac{\pi}{2}]$, $v \in (0, 1]$.

Through Examples 1 to 5, we demonstrate how to verify an isotropic variogram based on those necessary and/or sufficient conditions. Next we construct a type of polynomial variograms possessing a simple structure and hole effect.

Theorem 3. *Let* $\gamma(\vartheta)$ *be a real polynomial of degree* p *on* $[0, \pi]$,

$$\gamma(\vartheta) = \sum_{k=1}^{p} b_k \vartheta^k, \ \vartheta \in [0, \pi]. \tag{1.10}$$

(i) If the coefficients of $\gamma(\vartheta)$ *satisfy the following inequalities,*

$$(-1)^k \sum_{j=k}^{p} b_j \binom{j}{k} \left(\frac{\pi}{2}\right)^{j-k} \leq 0, \ k = 1, \ldots, p, \tag{1.11}$$

then $\gamma(\vartheta)$ *is an isotropic variogram on* $\mathbb{S}^\infty$.

(ii) If (1.10) is an isotropic variogram on $\mathbb{S}^\infty$, *then*

$$(-1)^k \sum_{j=k}^{p} b_j \binom{j}{k} \left(\frac{\pi}{2}\right)^{j-k} \leq 0, \ k = 1, 2. \tag{1.12}$$

(iii) When $p = 2$, *a quadratic polynomial*

$$\gamma(\vartheta) = b_1 \vartheta + b_2 \vartheta^2, \ \vartheta \in [0, \pi],$$

is an isotropic variogram on $\mathbb{S}^\infty$ *if and only if* $b_1 + b_2 \pi \geq 0$ *and* $b_2 \leq 0$.

It is known that an intrinsically stationary variogram $\gamma(x)$ on $\mathbb{R}$ possesses a subadditive property (see e.g. Schilling et al. 2012), which implies that $\gamma(x) \leq a_0 x^2, x \in \mathbb{R}$, for a

positive constant a_0. At the first sight, one might wonder if the polynomial structure (1.10) contradicts such a subadditive property. Fortunately, the answer is no, since one can deduce from (10) that $\gamma(\vartheta) \leq a_0\vartheta, \vartheta \in [0, \pi]$, where

$$a_0 = \max_{0\leq\vartheta\leq\pi} \left| \sum_{k=1}^{p} b_k \vartheta^{k-1} \right|.$$

The development of Theorem 3 allows us to fit a polynomial type of variograms on all spheres. In Section 5 we will show the advantage of this type of variograms in terms of capturing the negative spatial correlation for long distance kriging.

1.3 Multivariate Isotropic Variogram Models on All Spheres

This section deals with the matrix version of variogram models on all spheres, which account for not only the dependence of a random field at different locations, but also the interaction of multiple random fields observed on the same domain. An m-variate random field $\{\mathbf{Z}(\mathbf{x}) = (Z_1(\mathbf{x}), \ldots, Z_m(\mathbf{x}))', \mathbf{x} \in \mathbb{S}^d\}$ is said to be intrinsically stationary and isotropic, if the mean of $\mathbf{Z}(\mathbf{x}_1) - \mathbf{Z}(\mathbf{x}_2)$ does not depend on $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d$ and its variogram matrix function $\boldsymbol{\gamma}(\vartheta(\mathbf{x}_1, \mathbf{x}_2)) = (\boldsymbol{\gamma}_{ij}(\vartheta(\mathbf{x}_1, \mathbf{x}_2)))_{1\leq i,j\leq m}$ depends only on $\vartheta(\mathbf{x}_1, \mathbf{x}_2)$ between $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d$. The direct variogram $\gamma_{ii}(\vartheta(\mathbf{x}_1, \mathbf{x}_2))$ is the variogram of random field $\{Z_i(\mathbf{x}), \mathbf{x} \in \mathbb{S}^d\}$,

$$\gamma_{ii}(\vartheta(\mathbf{x}_1, \mathbf{x}_2)) = \frac{1}{2} \operatorname{var}(Z_i(\mathbf{x}_1) - Z_i(\mathbf{x}_2)), \ i = 1, \ldots, m. \tag{1.13}$$

The cross variogram reflexes the cross dependence between random fields $\{Z_i(\mathbf{x}), \mathbf{x} \in \mathbb{S}^d\}$ and $\{Z_j(\mathbf{x}), \mathbf{x} \in \mathbb{S}^d\}$ $(i \neq j)$,

$$\gamma_{ij}(\vartheta(\mathbf{x}_1, \mathbf{x}_2)) = \frac{1}{2}\operatorname{cov}(Z_i(\mathbf{x}_1) - Z_i(\mathbf{x}_2), Z_j(\mathbf{x}_1) - Z_j(\mathbf{x}_2)). \tag{1.14}$$

Similar to the scalar case, we say that $\gamma(\vartheta)$ is an isotropic variogram matrix function on all spheres if it is so for all $\mathbb{S}^d$ $(d = 1, 2, \ldots)$. An isotropic variogram matrix on all spheres is characterized in the following theorem.

Theorem 4. *For an $m \times m$ matrix function $\boldsymbol{\gamma}(\vartheta)$ whose entries are continuous on $[0, \pi]$, the following statements are equivalent:*

(i) $\boldsymbol{\gamma}(\vartheta)$ is an isotropic variogram matrix function on $\mathbb{S}^\infty$;

(ii) $\boldsymbol{\gamma}(\vartheta)$ is of the form

$$\boldsymbol{\gamma}(\vartheta) = \sum_{n=1}^{\infty} \mathbf{B}_n \left\{1 - \cos^n \vartheta\right\}, \ \vartheta \in [0, \pi], \tag{1.15}$$

where $\{\mathbf{B}_n, n = 1, 2, \ldots\}$ is a summable sequence of positive definite $m \times m$ matrices;

(iii) $\boldsymbol{\gamma}\left(\frac{\pi}{2} - \arcsin x\right)$ is of the form

$$\boldsymbol{\gamma}\left(\frac{\pi}{2} - \arcsin x\right) = \sum_{n=1}^{\infty} \mathbf{B}_n(1 - x^n), \ x \in [-1, 1], \tag{1.16}$$

where $\{\mathbf{B}_n, n = 1, 2, \ldots\}$ is a summable sequence of positive definite $m \times m$ matrices.

The next example is to extend theorem 3 part (iii) to multivariate case.

Example 6. *An $m \times m$ matrix function with quadratic polynomial entries*

$$\boldsymbol{\gamma}(\vartheta) = \mathbf{B}_1\vartheta + \mathbf{B}_2\vartheta^2,\ \vartheta \in [0, \pi],$$

is a variogram matrix function on $\mathbb{S}^\infty$ if and only if $\mathbf{B}_1 + \pi\mathbf{B}_2$ is a positive definite matrix and $\mathbf{B}_2$ is a negative definite matrix. To apply Theorem 4 (iii), we employ the Taylor series expansions of $\arcsin x$ *and* $(\arcsin x)^2$*,*

$$\arcsin x = \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2(2n+1)} x^{2n+1},\ x \in [-1, 1],$$

and

$$(\arcsin x)^2 = \sum_{n=1}^{\infty} \frac{2^{2n-1}((n-1)!)^2}{(2n)!} x^{2n},\ x \in [-1, 1],$$

to rewrite $\boldsymbol{\gamma}\left(\frac{\pi}{2} - \arcsin x\right)$ as follows:

$$\begin{aligned}
& \boldsymbol{\gamma}\left(\frac{\pi}{2} - \arcsin x\right) \\
= \ & \mathbf{B}_1\left(\frac{\pi}{2} - \arcsin x\right) + \mathbf{B}_2\left(\frac{\pi}{2} - \arcsin x\right)^2 \\
= \ & \frac{\pi}{2}\mathbf{B}_1 + \left(\frac{\pi}{2}\right)^2 \mathbf{B}_2 - (\mathbf{B}_1 + \pi\mathbf{B}_2)\arcsin x + \mathbf{B}_2(\arcsin x)^2 \\
= \ & \frac{\pi}{2}\mathbf{B}_1 + \left(\frac{\pi}{2}\right)^2 \mathbf{B}_2 - (\mathbf{B}_1 + \pi\mathbf{B}_2)\sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2(2n+1)} x^{2n+1} + \mathbf{B}_2 \sum_{n=1}^{\infty} \frac{2^{2n-1}((n-1)!)^2}{(2n)!} x^{2n} \\
= \ & (\mathbf{B}_1 + \pi\mathbf{B}_2)\sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2(2n+1)}(1 - x^{2n+1}) - \mathbf{B}_2 \sum_{n=1}^{\infty} \frac{2^{2n-1}((n-1)!)^2}{(2n)!}(1 - x^{2n}),\ x \in [-1, 1].
\end{aligned}$$

Clearly, the coefficient matrices in the last expansion are positive definite if and only if $\mathbf{B}_1 + \pi\mathbf{B}_2$ and $\mathbf{B}_2$ are positive definite.

Corollary 3. *For an $m \times m$ matrix function $\boldsymbol{\gamma}(\vartheta)$ whose entries are continuous on $[0, \pi]$,*

if there is a sequence of $m \times m$ *positive definite matrices* $\{\mathbf{B}_n, n = 0, 1, 2, \ldots\}$ *such that*

$$\boldsymbol{\gamma}\left(\frac{\pi}{2} - x\right) = \sum_{n=1}^{\infty} \mathbf{B}_n \left(\left(\frac{\pi}{2}\right)^n - x^n\right), \ x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \tag{1.17}$$

then $\boldsymbol{\gamma}(\vartheta)$ *is a variogram matrix function on* $\mathbb{S}^\infty$.

This corollary is confirmed once (1.16) is verified. To this end, we use the Taylor series of $(\arcsin x)^n$,

$$(\arcsin x)^n = \sum_{k=0}^{\infty} a_{kn} x^k, \ x \in [-1, 1], \ n = 1, 2, \ldots,$$

where $\{a_{kn}, k = 0, 1, \ldots\}$ is a summable sequence of nonnegative numbers for each n, and obtain

$$\begin{aligned}
\boldsymbol{\gamma}\left(\frac{\pi}{2} - \arcsin x\right) &= \sum_{n=1}^{\infty} \mathbf{B}_n \left[\left(\frac{\pi}{2}\right)^n - (\arcsin x)^n\right] \\
&= \sum_{n=1}^{\infty} \mathbf{B}_n \left[(\arcsin 1)^n - (\arcsin x)^n\right] \\
&= \sum_{n=1}^{\infty} \mathbf{B}_n \left[\sum_{k=0}^{\infty} a_{kn}(1 - x^k)\right] \\
&= \sum_{n=1}^{\infty} \left(\sum_{k=0}^{\infty} a_{kn} \mathbf{B}_k\right)(1 - x^n), \ x \in [-1, 1],
\end{aligned}$$

where the positive definiteness of $\sum_{k=0}^{\infty} a_{nk} \mathbf{B}_k$ is due to positive definiteness of $\mathbf{B}_k$ and non-negativeness of a_{kn} for each n.

The condition in Corollary 3 is sufficient for $\boldsymbol{\gamma}(\vartheta)$ to be an isotropic variogram matrix function on $\mathbb{S}^\infty$. An exponential type of variogram matrix function on S^∞ based on this corollary is derived in the next example.

Example 7. *An $m \times m$ matrix $\mathbf{B} = (b_{ij})$ being positive definite is a necessary and sufficient condition for the $m \times m$ functions*

$$\gamma_{ij}(\vartheta) = \exp\left(\frac{\pi}{2} b_{ij}\right)\left(1 - \exp(-b_{ij}\vartheta)\right),\ \vartheta \in [0, \pi], i, j = 1, \ldots, m, \tag{1.18}$$

to form an isotropic variogram matrix function on $\mathbb{S}^\infty$. This can be verified via Corollary 3 with the help of the Taylor expansion of $\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. In fact, for $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, plugging this expansion into (1.18) we obtain

$$\begin{aligned}
\gamma_{ij}\left(\frac{\pi}{2} - x\right) &= \exp\left(\frac{\pi}{2} b_{ij}\right) - \exp\left(\frac{\pi}{2} b_{ij}\right) \exp\left(-b_{ij}\left(\frac{\pi}{2} - x\right)\right) \\
&= \exp\left(\frac{\pi}{2} b_{ij}\right) - \exp(b_{ij} x) \\
&= \sum_{n=1}^{\infty} \frac{b_{ij}^n}{n!}\left(\left(\frac{\pi}{2}\right)^n - x^n\right),
\end{aligned}$$

where the coefficient matrices $\frac{1}{n!}\mathbf{B}^{\circ n}$ are positive definite if and only if $\mathbf{B}$ is so. Here $\mathbf{A} \circ \mathbf{B} = (a_{ij} b_{ij})_{m \times m}$ denotes entry-wise or Hadamard product of two matrices $\mathbf{A} = (a_{ij})_{m \times m}$ and $\mathbf{B} = (b_{ij})_{m \times m}$, and $\mathbf{B}^{\circ n}$ is the nth Hadamard power of $\mathbf{B}$.

Example 8. *For $\lambda \in (0, 1]$, ϑ^λ is a variogram on $\mathbb{S}^\infty$, as is shown in Du et al. (2013). With an $m \times m$ positive definite matrix $\mathbf{B}_0$, $\mathbf{B}_0 \vartheta^\lambda$ is a variogram matrix function on $\mathbb{S}^\infty$, by Lemma 2(i) of Du and Ma (2012). So is*

$$\gamma(\vartheta) = \mathbf{B}_0 \vartheta^\lambda + \mathbf{B}_1 \vartheta + \mathbf{B}_2 \vartheta^2,\ \vartheta \in [0, \pi], \tag{1.19}$$

as the summation of $\mathbf{B}_0 \vartheta^\lambda$ and $\mathbf{B}_1 \vartheta + \mathbf{B}_2 \vartheta^2$, provided that $2\mathbf{B}_1 + \pi \mathbf{B}_2, \mathbf{B}_1 + \mathbf{B}_2 \pi$ are positive definite and $-\mathbf{B}_2$ are negative definite (See Example 6).

The variogram matrix functions in Examples 6 and 8 may be lack of flexibility when fitting a real data set, but they may have a better performance for some intrinsic stationary processes. Example 7 is used in one of the simulation study in the next section to demonstrate the advantage of our model.

1.4 Simulation Studies

Common random field simulation approaches in literature are typically conducted through defining a valid covariance (matrix) first and the resulting field is often limited to be stationary and Gaussian. In this section, we propose a new method to simulate scalar or vector random fields on a sphere $\mathbb{S}^d$ $(d \geq 2)$ based on a series expansion of a random field described in the following theorem, which is not necessarily Gaussian. The simulated process will be used to assess models developed in previous sections.

For a positive constant λ, the ultraspherical or Gegenbauer's polynomials $P_n^{(\lambda)}(x)$ $(n = 0, 1, 2, \ldots)$ can be defined through the following recurring formula

$$\begin{aligned} P_0^{(\lambda)}(x) &= 1, \\ P_1^{(\lambda)}(x) &= 2\lambda x, \\ P_n^{(\lambda)}(x) &= \frac{2(\lambda+n-1)xP_{n-1}^{(\lambda)}(x) - (2\lambda+n-2)P_{n-2}^{(\lambda)}(x)}{n}, \ x \in \mathbb{R}, n \geq 2. \end{aligned}$$

We further define $p_n^{\left(\frac{d-1}{2}\right)}(x) = \frac{P_n^{\left(\frac{d-1}{2}\right)}(x)}{P_n^{\left(\frac{d-1}{2}\right)}(1)}$ as in Du et al. (2013). Let $\{\alpha_n, n \in \mathbb{N}\}$ be a sequence of positive numbers defined by

$$\alpha_n = \left(\frac{2n+d-1}{d-1}\right)^{\frac{1}{2}}, n \in \mathbb{N}. \tag{1.20}$$

For a positive definite matrix $\mathbf{B}$, denote its square root by $\mathbf{B}^{\frac{1}{2}}$, which is a matrix of the same order of $\mathbf{B}$ such that $\mathbf{B} = \mathbf{B}^{\frac{1}{2}} \left(\mathbf{B}^{\frac{1}{2}}\right)'$. An $m \times m$ identity matrix is denoted by $\mathbf{I_m}$.

Theorem 5. *Assume that* $\{\mathbf{V}_n, n \in \mathbb{N}\}$ *is a sequence of independent* m*-variate random vectors with* $\mathrm{E}\mathbf{V}_n = 0$ *and* $\mathrm{cov}(\mathbf{V}_n, \mathbf{V}_n) = \alpha_n^2 \mathbf{I}_m$, $\mathbf{U}$ *is a (d+1)-dimensional random vector uniformly distributed on* $\mathbb{S}^d$ *and is independent of* $\{\mathbf{V}_n, n \in \mathbb{N}\}$, $\mathbf{Z}_0$ *is an* m*-variate random vector, and* $\{\mathbf{B}_n, n \in \mathbb{N}\}$ *is a sequence of* $m \times m$ *positive definite matrices. If the series* $\sum_{n=1}^{\infty} \mathbf{B}_n$ *converges, then*

$$\mathbf{Z}(\mathbf{x}) = \sum_{n=1}^{\infty} \mathbf{B}_n^{\frac{1}{2}} \mathbf{V}_n \left(\alpha_n^{-1} - \frac{P_n^{\left(\frac{d-1}{2}\right)}(\mathbf{x}'\mathbf{U})}{\left(P_n^{\left(\frac{d-1}{2}\right)}(1)\right)^{\frac{1}{2}}} \right) + \mathbf{Z}_0, \ \mathbf{x} \in \mathbb{S}^d, \tag{1.21}$$

is an m*-variate random field on* $\mathbb{S}^d$*, with the isotropic variogram matrix function*

$$\boldsymbol{\gamma}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{n=1}^{\infty} \mathbf{B}_n \left\{ 1 - p_n^{\left(\frac{d-1}{2}\right)}(\cos(\vartheta(\mathbf{x}_1, \mathbf{x}_2))) \right\}, \ \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d. \tag{1.22}$$

Expression (1.22) turns out to be a general form of isotropic variogram matrix functions on $\mathbb{S}^d$, according to Theorem 1 of Du et al. (2013). Theorem 5 enables us to construct an intrinsic random field by simply simulating an independent sequence $\mathbf{V}_n$, and uniform distributed random vector $\mathbf{U}$ on $\mathbb{S}^d$, while the resulting random field may or may not be Gaussian. If $\mathbf{Z}_0$ does not have a second order moment, then (1.21) is not a second order random field, but its increment is so. A random vector $\mathbf{U}$ uniformly distributed on $\mathbb{S}^d$ can be formulated by $\mathbf{U} = \left(\frac{V_1}{\left(\sum_{k=1}^{d+1} V_k^2\right)^{\frac{1}{2}}}, \ldots, \frac{V_{d+1}}{\left(\sum_{k=1}^{d+1} V_k^2\right)^{\frac{1}{2}}} \right)'$, where $V_1, \ldots, V_{d+1}$ are independent and standard normal random variables. One way to obtain a sequence of positive definite matrices $\mathbf{B}_n$ is to repeat Hadamard product of a positive definite matrix. In simulation

study, we always want to make sure the condition for model comparison is fair for all models in check. Knowing the explicit form of covariance structure for a random field to simulate from, may inevitably allow a researcher to prefer one model over another. By Theorem 5, we attempt to remove this selection bias during the simulation study.

1.4.1 Simulation I

In the previous section, we develop a theory of building variogram models on all spheres. Jeong and Jun (2015) argues the need to construct variogram specifically built for spheres, even when matérn or other variogram can be valid on all spheres through this $2sin(\frac{\theta}{2})$ transformation. It is pointed out that correlation function exists in $\mathbb{R}^3$ cannot have value less than -0.218. Matérn class is by far the most popular models for fitting spatial data on Euclidean space due to its great flexibility and theoretical justification. Actually, matérn class possesses a smoothness parameter, which governs differentiability of covariance function at origin, as a result it is capable of predicting well if there are many points near the predicted locations.

In our simulation, suppose the data is generated from covariance function with hole effect, which can generate negative covariance. One hundred locations are randomly generated on unit spheres. In order to make data more spread out, we follow Jeong and Jun (2015)'s approach to sample the locations with percentages depending on the latitude, roughly proportional to the area of each latitude band. The final number distribution of points is $(26, 24, 21, 16, 10, 3)$ for latitude $(\pm[0, 15], \pm(15, 30], \pm(30, 45], \pm(45, 60], \pm(60, 75], \pm(75, 90])$ degree. In this simulation study, we simulate mean zero Gaussian random field on $\mathbb{S}^2$ with the following covariance structure.

$$C(t) = \sigma^2 \frac{\alpha}{t} sin(\frac{t}{\alpha}), t > 0$$

where $C(0) = \sigma^2$, which is set to be 1 and α is set to be 0.2. By Theorem 4 in Du et al. (2013), the hole effect model is valid on $\mathbb{S}^2$ by letting $t = 2sin(\vartheta/2)$. Also, we set $\alpha = 0.2$.

We compare polynomial of order 2 variogram function (P2) developed in theorem 3(iii) with Matérn variogram (MG), which is valid on $\mathbb{S}^2$ only when the smoothness parameter is restrict to $[0, 0.5]$. Under chordal distance with transformation $t = 2sin(\vartheta/2)$, a Matérn variogram is fitted as the third comparison (MC). Furthermore, power variogram (PL) is also introduced to this comparison, which has a form of

$$\gamma(\vartheta) = b_0\vartheta^{\lambda}, \lambda \in [0, 1], b_0 > 0$$

.

Among 100 data points, we use 90 randomly selected data points as training data and remaining 10 points for testing. The Cressie's weighted least square is used for parameter estimation. 100 replications of simulations are run to compare this four models. Table (1.1) displays mean and standard deviations for Root Mean Square Error (RMSE) and Continuous Ranked Probability Score (CRPS) of the prediction (See e.g. Gneiting and Raftery (2007)). From Table 1, we observe Polynomial of order 2 model performs the best with smallest RMSE and CRPS among the competitive models Matérn on spherical distance, Matérn on chordal distance and power variogram model. Figure 1.1 shows an example of empirical variogram along with the fit from those four models. It is clear that polynomial type of variogram's shape is model honoring that of empirical variogram even in the near zero area. Note that the Hole effect model has negative covariance and the realized data is often spread out across the sphere. We might expect the polynomial type of model would work well when those conditions are met.

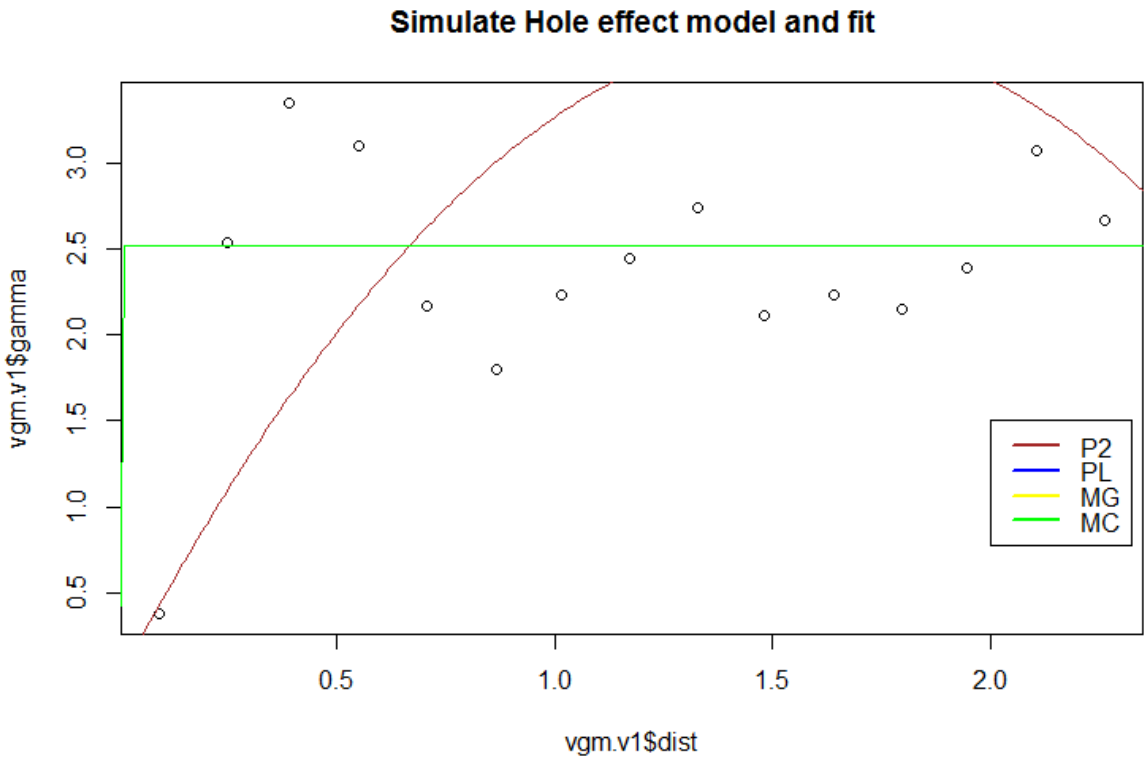

Figure 1.1: *The empirical variogram of hole-effect model simulated and four types of variogram models' fit*

1.4.2 Simulation II

The cokriging procedure is a natural extension of kriging when multivariate spatial data is concerned (Cressie 1993). Often times in a multivariate scenario, we are interested in prediction of a primary variable, which has limited sample, then correlation with multiple auxiliary variables can be borrowed to improve prediction. Namely a variable of interest can be cokriged at certain locations through data of itself and auxiliary variables nearby. As in (Du et al. 2013), cokriging performance of some proposed models from previous section will be evaluated in comparison with popular multivariate techniques

Measure	P2	MG	MC	PL
RMSE(SD)	1.133(0.327)	1.317(0.328)	1.255(0.343)	1.233(0.359)
CRPS(SD)	0.667(0.219)	0.766(0.193)	0.726(0.203)	0.710(0.215)

Table 1.1: *RMSE and CRPS of four types of model in simulation I*

such as linear model of coregionalization (LMC). LMC assumes that all direct and cross variogram structures are linear combinations of the same uncorrelated basis components observed at different spatial scales (IM et al. 2007). A exponential variogram with a partial sill=1.1 and range parameter =5000 is selected to be the basic component of LMC in this simulation study.

In this simulation, the data is generated through a bivariate Matérn variogram and the first variable is the main variable. The direct variogram and cross-variogram have a form of

$$\begin{aligned}\gamma_{ii}(\vartheta) &= \sigma^2(1 - M(2sin(\frac{\vartheta}{2}|\nu_i, \alpha_i))) \; i = 1, 2, \\ \gamma_{ij}(\vartheta) &= \rho_{ij}\sigma_i\sigma_j(1 - M(2sin(\frac{\vartheta}{2}|\nu_{ij}, \alpha_{ij}))) \; i = 1, j = 2,\end{aligned}$$

where $M(\cdot|\nu, \alpha)$ is a Matérn correlation function with smooth parameter ν and scale parameter α. The true value of parameters is $\nu_1 = \nu_2 = \nu_{12} = 1.5, \sigma_1 = \sigma_2 = 1, \alpha_1 = 2, \alpha_2 = 2.5, \alpha_{12} = 3$ and $\rho_{12} = 0.185$ based on the equation (12) and (16) in Gneiting et al. (2010).

The first coordinate variable in (1.21) is set to be the primary variable for co-kriging. In 100 replications of simulation, we randomly select 500 points on unit sphere each time as sampled location. We choose 20 fixed locations in a regular grid for prediction with latitude in $\{-\frac{\pi}{3}, -\frac{\pi}{6}, 0, \frac{\pi}{6}, \frac{\pi}{3}\}$ and longtitute in $\{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ on unit sphere.

We will demonstrate the performance of the exponential type multivariate variogram

Locations (Lon, Lat)	LMC IPE	ME IPE	Locations (Lon, Lat)	LMC IPE	ME IPE
$(0, -\frac{\pi}{3})$	0.00576	0.00264	$(\pi, 0)$	0.36335	0.00415
$(\frac{\pi}{2}, -\frac{\pi}{3})$	0.02272	0.00264	$(\frac{3\pi}{2}, 0)$	0.01578	0.00367
$(\pi, -\frac{\pi}{3})$	0.01006	0.00345	$(0, \frac{\pi}{6})$	0.0604	0.00339
$(\frac{3\pi}{2}, -\frac{\pi}{3})$	0.00278	0.00205	$(\frac{\pi}{2}, \frac{\pi}{6})$	0.00929	0.00191
$(0, -\frac{\pi}{6})$	0.04502	0.00368	$(\pi, \frac{\pi}{6})$	0.01847	0.00281
$(\frac{\pi}{2}, -\frac{\pi}{6})$	0.01568	0.00300	$(\frac{3\pi}{2}, \frac{\pi}{6})$	0.04892	0.00283
$(\pi, -\frac{\pi}{6})$	0.01807	0.00373	$(0, \frac{\pi}{3})$	0.05310	0.00145
$(\frac{3\pi}{2}, -\frac{\pi}{6})$	0.04956	0.00399	$(\frac{\pi}{2}, \frac{\pi}{3})$	0.01011	0.00285
$(0, 0)$	0.11012	0.00306	$(\pi, \frac{\pi}{3})$	0.06319	0.00370
$(\frac{\pi}{2}, 0)$	0.56037	0.00151	$(\frac{3\pi}{2}, \frac{\pi}{3})$	0.01803	0.00304

Table 1.2: *Cokriged IPE for LMC and ME in simulation II*

matrix function (ME) proposed in Example 7 by comparing with the LMC in terms of Increase in Prediction Error (IPE), which is estimated by $\frac{1}{N}\sum_{i=1}^{N}(\hat{Z}_{1i}(\mathbf{x}_0) - \hat{Z}_{1i}^{true}(\mathbf{x}_0))^2$. Where N is the number of simulations, $\hat{Z}_1^{true}(\mathbf{x}_0)$ is the predicted value at location $\mathbf{x}_0$ using the true variogram matrix function for co-kriging, and $\hat{Z}_1(\mathbf{x}_0)$ is obtained using the estimated variogram function (IM et al. 2007). Simply speaking, IPE measures the extra mean squared prediction error introduced by using an estimated variogram rather than the true one. Obviously the model with smaller IPE is preferred.

At each of 20 locations, we compare the IPE of the ME model with LMC for in cokriging capability for this simulate dataset. Table (1.2) shows that the ME model proposed performs uniformly better than the LMC at all locations. At some of the locations, the ME's IPE value is only as small as 10% of that of LMC.

1.4.3 Simulation III

Simulation I discusses univariate variogram and the multivariate extension is discussed in simulation II. In this subsection We illustrate the application of theorem 5, which pro-

Locations (Lon, Lat)	LMC IPE	ME IPE	Locations (Lon, Lat)	LMC IPE	ME IPE
$(0, -\frac{\pi}{3})$	0.01903	0.00363	$(\pi, 0)$	0.01985	0.00463
$(\frac{\pi}{2}, -\frac{\pi}{3})$	0.01477	0.00256	$(\frac{3\pi}{2}, 0)$	0.03689	0.00219
$(\pi, -\frac{\pi}{3})$	0.00798	0.00274	$(0, \frac{\pi}{6})$	0.02271	0.00257
$(\frac{3\pi}{2}, -\frac{\pi}{3})$	0.00464	0.00215	$(\frac{\pi}{2}, \frac{\pi}{6})$	0.01759	0.00369
$(0, -\frac{\pi}{6})$	0.03675	0.00342	$(\pi, \frac{\pi}{6})$	0.01453	0.00233
$(\frac{\pi}{2}, -\frac{\pi}{6})$	0.02953	0.00274	$(\frac{3\pi}{2}, \frac{\pi}{6})$	0.00453	0.00235
$(\pi, -\frac{\pi}{6})$	0.01082	0.00333	$(0, \frac{\pi}{3})$	0.0393	0.00215
$(\frac{3\pi}{2}, -\frac{\pi}{6})$	0.02225	0.00422	$(\frac{\pi}{2}, \frac{\pi}{3})$	0.02229	0.00657
$(0, 0)$	0.12779	0.00266	$(\pi, \frac{\pi}{3})$	0.01491	0.00215
$(\frac{\pi}{2}, 0)$	0.00228	0.00275	$(\frac{3\pi}{2}, \frac{\pi}{3})$	0.01559	0.00186

Table 1.3: *Cokriged IPE for LMC and ME in simulation III*

vides a method to simulate random field on the unit sphere directly from a stochastic series. Through this approach, we can simulate non-stationary or non-gaussian process. This method also allow a relatively fair comparison on the performance of exiting theoretical variogram and variogram proposed by this paper with less selection bias when no covariance structure is specified in simulating from this constructed random field.

In our simulation study, we choose $\mathbf{V}_n$ as a sequence of bi-variate normal random vectors with $\mathrm{E}\mathbf{V}_n = 0$ and $\mathrm{cov}(\mathbf{V}_n, \mathbf{V}_n) = \alpha_n^2 \mathbf{I}_2$. According to equation (1.20), $\alpha_n = (2n+1)^{\frac{1}{2}}$, $n \in \mathbb{N}$ with dimension $d = 2$. Let $\mathbf{U}$ be a random vector that is uniformly distributed on $\mathbb{S}^2$ and $\mathbf{B}_n = {\mathbf{B}_0}^{\circ n}$, $n = 0, 1, 2, \ldots$ with a positive definite matrix $\mathbf{B}_0 = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{2}{3} \end{bmatrix}$.

We randomly sample 500 locations from the random field specified earlier based on Theorem 5. Twenty fixed locations in a regular grid are chosen for prediction with latitude in $\{-\frac{\pi}{3}, -\frac{\pi}{6}, 0, \frac{\pi}{6}, \frac{\pi}{3}\}$ and longtitute in $\{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ on a unit sphere. At each of 20 locations, the estimated increase in prediction error based on both the ME model and

LMC are computed. Table 1.1 shows that the ME model performs uniformly better than the LMC at all locations in IPE. At some of the locations, the ME's IPE error is only 10% of LMC's. This superiority over LMC has been found no matter how the data is generated, as long as the empirical variogram roughly takes the shape of the theoretical model proposed.

1.5 Data Analysis

With the advancement of techniques in data collection, the need to study large scale data sets has received increasing attention. In order to demonstrate the application and performance of polynomial-type variograms developed in section 2, we use the geopotential height data set studied in Jeong and Jun (2015) and compare the model proposed in Theorem 3 with those models selected in Jeong and Jun (2015). This global data set originates from NCEP/NCAR reanalysis project (http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html), which is using a state-of-the-art analysis/forecast system to perform data assimilation using past data from 1948 to the present. A subset of this data from June to August 2014 at 500 hPa level is used and located on regular grids with 144 longitude points and 73 latitude points. Geopotential height approximates the actual height of certain pressure surface above mean sea level. It has been shown the geopotential height at 500 hPa is a dominant parameter in controlling weather and climate condition over the globe (Hafez and Almazroui 2015).

Following Jeong and Jun (2015), we remove the temporal effect by averaging over three months of data on each location and stabilize the variance through a square root transformation of the data. The modeling of the resulting data consists of mean structure specification and covariance construction for the residual field after the mean is removed.

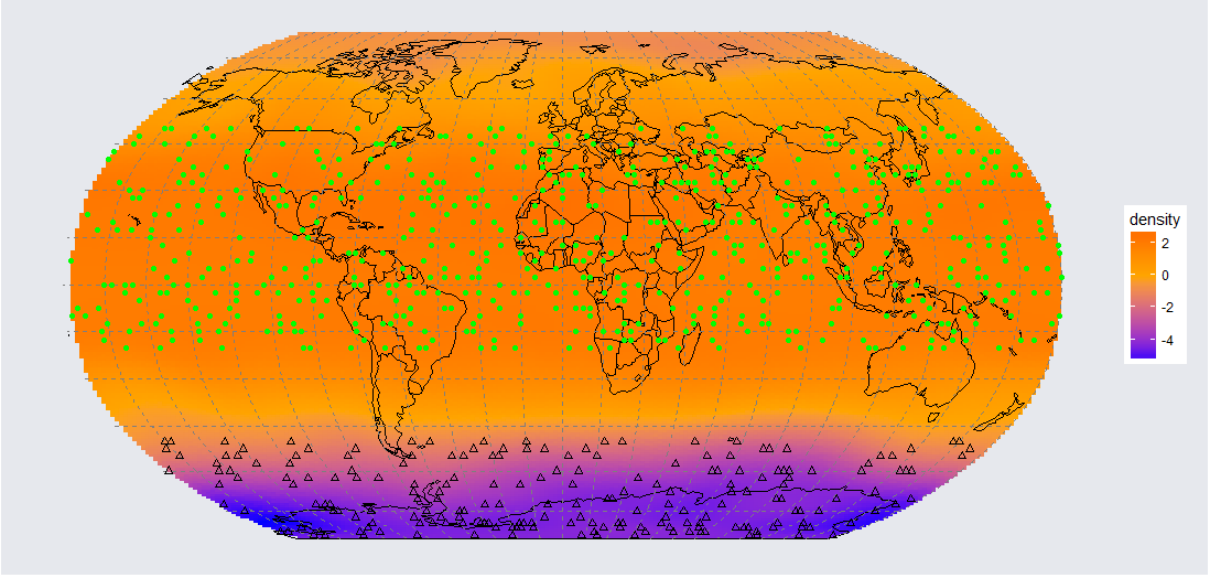

Figure 1.2: *Image plot of the residual after removing overall mean. The green dots are sampled locations in one replication and the triangular is test site for this replication*

Two methods are used to fit the mean structure: method 1 uses overall mean, method 2 is based on harmonic regression in latitude L as in Jeong and Jun (2015) with mean

$$\mu(L) = \beta_0 + \beta_1 \cos(\frac{\pi L}{90}) + \beta_2 \sin(\frac{\pi L}{90}). \tag{1.23}$$

Figure 1.2 gives an image plot of the residual after method 1 is applied. The proposed models in Theorem 3 will be assessed in terms of prediction by popular criteria of Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Continuous Ranked Probability Score (CRPS) (Gneiting and Raftery 2007) for the residual random field. To this end, we repeat the following process 100 times. Each time we randomly sample 600 training sites with latitude between $-20°$ and $+50°$ and randomly select 200 testing sites with latitude between $-90°$ and $-50°$. Training data is used to predict at testing sites and the prediction criteria values are averaged over all simulations. Three competitive

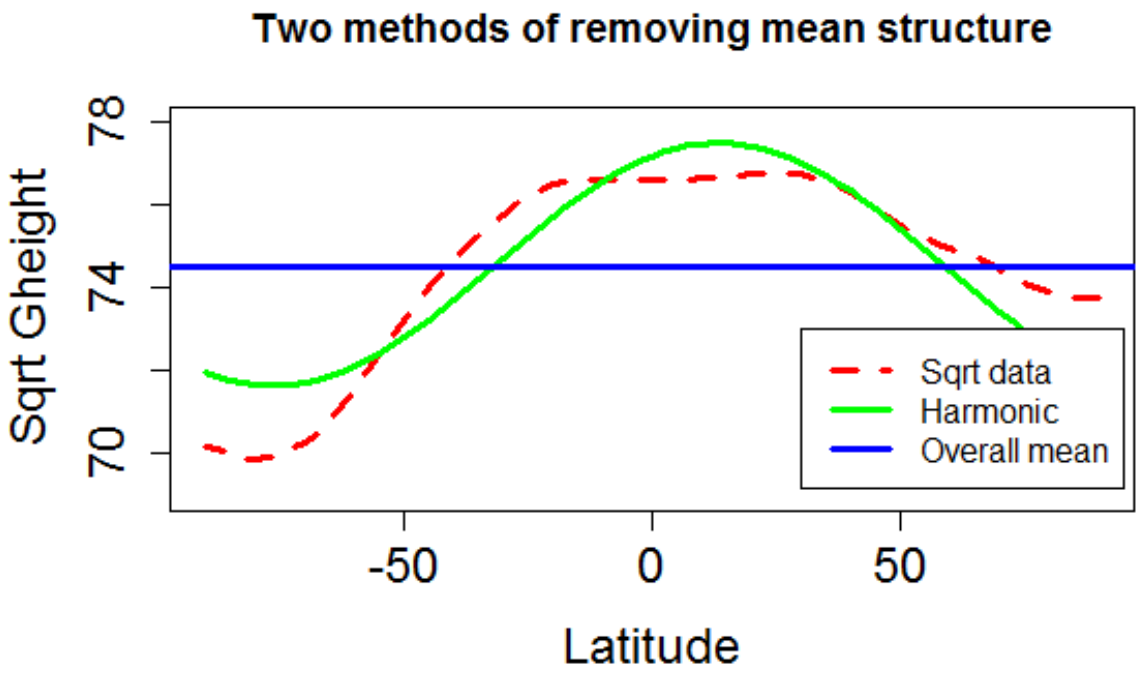

Figure 1.3: *Distribution of square root geopotential height over latitude with fits in Harmonic regression*

models adopted in Jeong and Jun (2015) are used to compare with quadratic or polynomial of order 2 model (P2) developed in Theorem 3 (iii). They are Matérn in spherical distance (MG), Matérn in Euclidean distance (ME) and a convex sum of sine-power and hole-effect model (C) defined in Jeong and Jun (2015). As we know, Matérn class of variograms is by far the most popular model for fitting spatial data on Euclidean space due to its great flexibility and theoretical justification (Stein 1999). The Matérn class possesses a smoothness parameter, which governs differentiability of covariance function at the origin, as a result it is normally capable of predicting well if there are sufficient points near the predicting locations. However this type of model is valid on $\mathbb{S}^2$ only when smoothness parameter falls in $(0, 0.5]$ (Gneiting 2013). This model is identified as MG

here. The model ME is referred to Matérn model on $\mathbb{R}^3$ restricted to $\mathbb{S}^2$ by the mapping in equation (1.1), which inherits its differentiability properties and therefore its powerful capability of interpolating well in general as shown in Guinness and Fuentes (2016). The form of C model is converted to a variogram form as follows:

$$\gamma(\vartheta) = \sigma^2 - \sigma^2\{\lambda[1 - (\sin\vartheta)^\beta] + (1-\lambda)\cos\vartheta\},\ \beta \in (0,2] \text{ and } \lambda \in (0,1) \tag{1.24}$$

Table 1.4 contains the results of the four models above in terms of RMSE, MAE and CRPS. In this sampling scheme, there is a gap between the training sites and testing sites in about 30° degree of angle. Simulation study suggests that our polynomial type of model has its advantage when data is relatively sparse on the sphere. This sampling scheme will enable us to check how well the variogram fits when spherical distance is larger than $\frac{\pi}{6}$ on a unit sphere. In Figure 1.4, we observe that the quadratic variogram behaves differently from MC, MG and C model by bending a little bit when spherical distance is larger than 1.5. Actually we can see that the proposed model outperforms the rest of the models for this long-distance prediction as suggested in Table 1.2, which shows that proposed (P2) model has smallest value in all RMSE, MAE and CRPS. One may argue that the decreasing trend of the data from north to south hasn't been adjusted and it may cause empirical variogram ill-behaved. Next, by method 2 we place a more complex structure on the mean and conduct model comparison in terms of prediction using the residual field.

Figure 1.3 shows the harmonic regression fit for mean structure through equation 1.23. The trend is removed before we fit our varogram models. Figure 1.5 shows the image plot of the residual field after the trend is fit using simple harmonic regression in latitude similar to that in Jeong and Jun (2015) and the large scale trend is removed. As

Measure	P2	MG	MC	C
RMSE	4.058	5.707	5.444	5.690
MAE	3.943	5.584	5.317	5.567
CRPS	3.832	5.434	5.164	5.425

Table 1.4: *Data analysis results with constant mean removed*

Measure	P3	MG	MC	C	W
RMSE	1.329	1.785	1.871	1.864	1.420
MAE	1.195	1.603	1.688	1.680	1.264
CRPS	1.002	1.369	1.450	1.442	1.033

Table 1.5: *Model comparisons after subtracting harmonic regression mean*

in the case based on constant mean assumption by method 1, we still randomly select 600 training sites among latitude from $-20°$ to $+50°$ and sample 200 testing sites among latitude from $-90°$ to $-50°$. Again we predict based on training data and repeat for 100 times. We compare the proposed polynomial of order 3 variogram (P3) in Theorem 3 with those models used in Jeong and Jun (2015) again. Those are Matérn in Euclidean distance (MC), Matérn in spherical distance (MG), convex sum model in equation 1.24 (C) and C^4-Wendland (W), which is a compactly supported variogram. The reason of fitting a cubic polynomial instead of quadratic polynomial is suggested by the waved shape of empirical variogram. In figure 1.6, empirical variogram goes up fast from 0 to 0.8, then level down until it goes up again after 1.7. The results are revealed in Table 1.5, which shows that cubic model (P3) on average has smallest value in all RMSE, MAE and CRPS comparing with the other competitive models MG, MC, C and W. Again this polynomial type variogram model has its advantage of capturing the irregularly waved structure and has better capability in distant extrapolation.

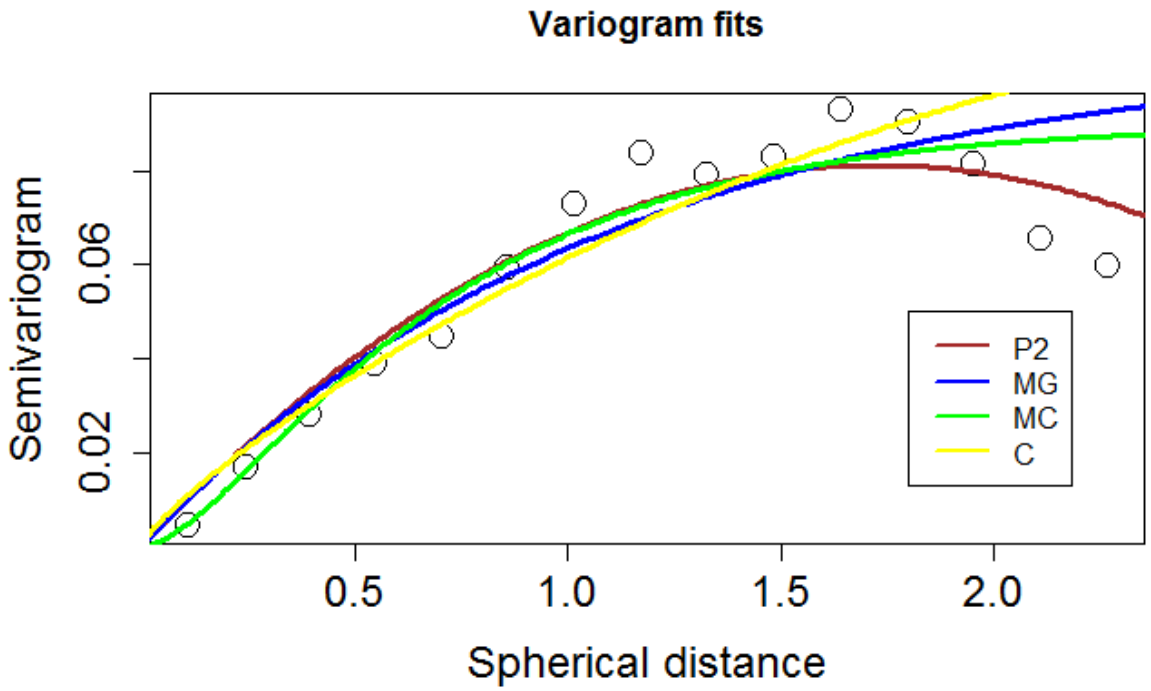

Figure 1.4: *Empirical variogram fits after removing overall mean*

1.6 Conclusion Remarks

While only unit spheres are mainly addressed in this paper, it can be shown that all theorems work well for an arbitrary radius of a sphere. When the radius is r, the series expansion (5) holds with $\mathbf{U}$ replaced by a (d+1)-variate random vector uniformly distributed on a sphere with radius $\frac{1}{r}$. We characterize the isotropic variogram models on all spheres by giving the sufficient and/or necessary conditions. These conditions can then be used to find new form of variograms on all spheres or to determine if an existing variogram is valid on all spheres. Some efficient approaches are proposed to construct a variety of variogram models including simple polynomial models. To identify the best order for a polynomial model, we recommend first take a look at the empirical variogram.

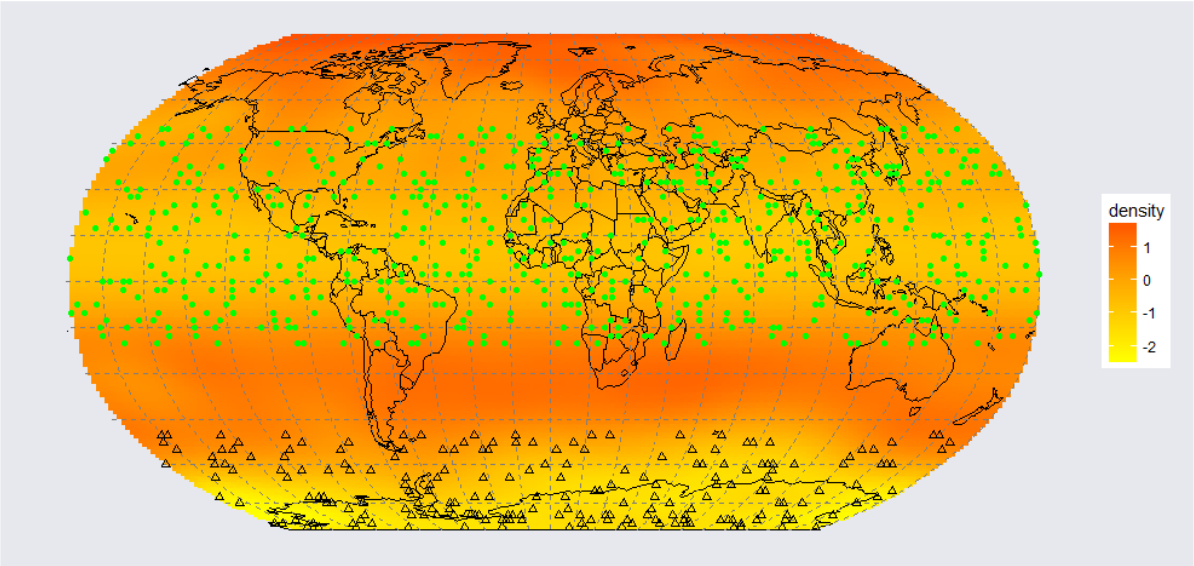

Figure 1.5: *Image plot of residual after subtracting harmonic regression mean. The green dots are sampled locations in one replication and the triangular is test site for this replication*

If the empirical variogram goes down significantly after reaching a certain spherical distance, one should consider quadratic polynomial. If not, a polynomial of order 3 could have a better fit than order 2. When order increases to 4 or more, we generally find that improvement through higher order is very small. Through the simulation study and data analysis, we find that the proposed polynomial type of models can make a considerable improvement in prediction when a distant forecasting or extrapolation on a sphere is considered. Though they do not have a good smoothness property near origin, which limits its ability in near distance interpolation. In addition, a series representation of multivariate random field on all spheres provides an effective way to simulate intrinsic stationary spatial processes without presetting any analytical variogram expression. In this work, we derive this representation and demonstrate its flexibility in comparing performance of different variogram models.

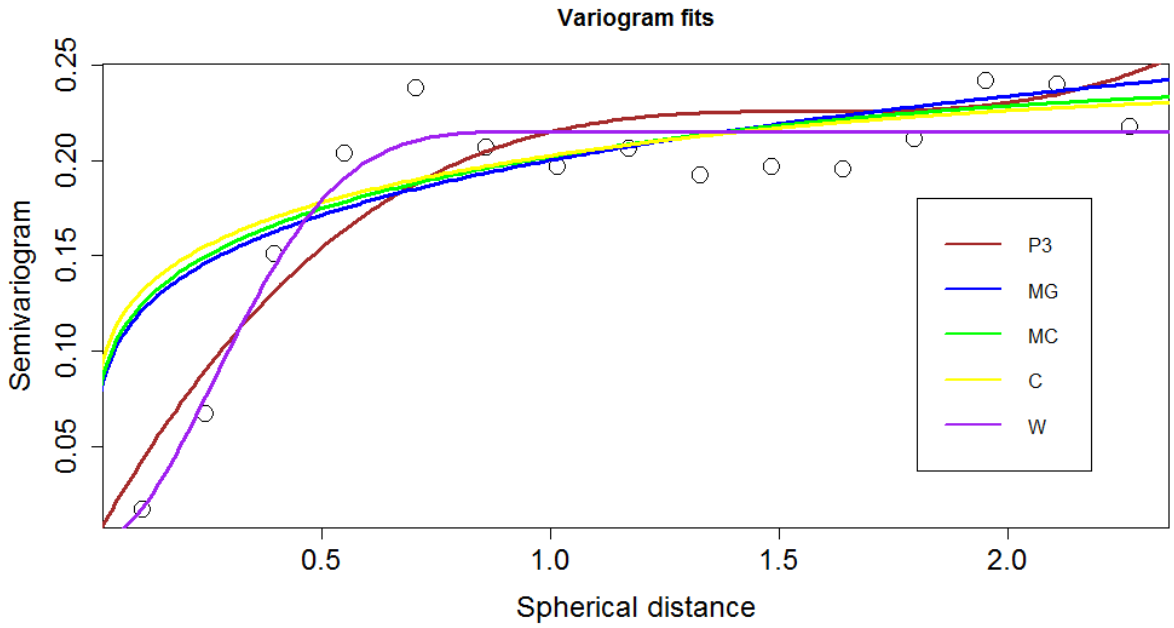

Figure 1.6: *Empirical variogram fits after subtracting harmonic regression mean*

Chapter 2

Dynamic Modeling of Days Suitable for Fieldwork Relative to Global Climate Cycles using INLA

2.1 Introduction and motivation

In an effort to achieve high yields with low cost, all producers strive to plant and harvest crops during optimal planting windows. When conducting field of operations, such as tillage, planting, spraying and harvesting, time is important manner to maximize the profit margin (Mark et al. 2014a). Days suitable for fieldwork (DSFW) is the count of work days for farming operation within a week or a month. The rationale for being concerned with DSFW is that during weather events that shorten DSFW, increased capacity from larger equipment or additional units of equipment are needed to achieve a timely planting or harvest of crops for a given number of acres (Mark et al. 2014a). This information

is different with knowing wet or dry in the next day. Locally, DSFW gives famers an expectation on how many days they could work in the field within a week or a month. They can purchase more or less equipment or arrange people to join the farming operations based on expectation of being under equipped with respect to planting or harvesting machinary (Griffin and Kelley 2010).

Predicting the number days suitable for fieldwork (DSFW) and its implications on agricultural production has been a topic of discussion in agricultural literature for more than five decades (Mark et al. 2014b). Also, climate change poses a major challenge to U.S. agriculture because of the critical dependence of the agricultural system on climate and because of the complex role agriculture plays in rural and national social and economic systems (Hatfield et al. 2014). However, the relationship between the climate change and DSFW remains unexplored in the literature. To be specific, this chapter investigates the influence of global Climate Cycles such as El Niño and La Niña on DSFW.

The El Niño Southern Oscillation (ENSO) is a naturally occurring phenomenon that involves fluctuation of ocean temperatures in the equatorial Pacific. It has three stages, El Niño, neutral and La Niña. When the sea surface temperatures are warmer than normal central and eastern equatorial, El Niño happens. When the sea surface temperatures are cooler than normal central and eastern equatorial, La Niña happens (Philander 1989). One of the measure for ENSO is the Oceanic Niño Index (ONI), which is a standardized index to observe periodicity of ENSO (Bamston et al. 1997). While impacts of ENSO are well studied on various of subjects, a closer look at how ENSO changes farmer's practice is still significant to agricultural decision making.

Being able to predict DSFW and know if the global circulation phenomenon such as ENSO's impact on the DSFW has a real benefit from farmer's day-to-day management decision to the state level policy changes . Ultimately, solving this question will improve

the efficiency all agriculture practice not only at a local farm scale but also at the state level (Mark et al. (2014a,b, 2015)).

In most of the spatial/spatio-temporal problems, the regression coefficients are considered to be fixed over time and space. In certain applications, this may not be appropriate. The coefficients may be expected to vary at different time or at different local or subregion level (Gelfand et al. (2003)). For instance, ENSO may have a complex effect on the DSFW. The way to define an El Niño events is when ONI has a high value continuously over 3 month and the ONI each month is computed over a standardized temperature changes over Niño 3.4 region(). The effect of DSFW would be different over time or over both time and space. Therefore, if we assume the effect of ENSO does not change over time, it is possible that the true effect may be cancelling out each other over time and space and yield an insignificant effect overall.

State-space models, also known as dynamic models in the Bayesian literature, are broad class of parametric models with time varying parameters where both, the parameter variation and the data information are described in a probabilistic way (Ruiz-Cárdenas et al. (2012)). According to Migon et al. (2005), dynamic models can be seen as a generalization of regression models, allowing changes in parameter values throughout time by the introduction of an equation governing the temporal evolution of regression coefficients. In this research, we propose various type of dynamic models to illustrate the complex effect of ENSO on DSFW over time and space through a computational framework known as Integrated Nested Laplace Approximation (INLA).

INLA is proposed by Rue and Martino (2007) and Rue et al. (2009) to perform fast full Bayesian inference through accurate approximation of marginal posterior and latent variables for a subclass of structured additive model, named Latent Gaussian model. This class of statistical models accommodate wide variety of models as long as the response is

within exponential family. Regression (generalized regression) models, dynamic models and spatial / spatio-temporal models are all belong to this class.

The reason to adopt INLA instead of MCMC in Bayesian estimation is because of its computational reduction over MCMC. It's obvious an approximated inference can reduce computational time by magnitude comparing with a sample algorithm such as MCMC. Moreover, MCMC's error can be made aribitrarily sma;; for arbitrarily large computational time. We also argue that INLA outperform MCMC algorithm for a given computational cost. This is a big deal especially in the spatio-temporal models due to the fact that the dimensionality in those models can be big but the computational cost for inverting a matrix is in the scale of $\mathcal{O}(n^3)$. So, the computational time for the dynamic models we fit may decrease from days in MCMC to hours in INLA.

However, direct fitting a dynamic model in INLA is not straight-forward. Even through, the INLA library under R has various latent models options and examples, fitting a flexible dynamic models tailored to the data at hand could be a problem. Ruiz-Cárdenas et al. (2012) provides some examples to fit dynamics models in INLA using a data augmentation approach. However, it assumes the underlying spatial domain and temporal domain is both discrete. Cameletti et al. (2013) introduces an approach for continuous spatial domain and a discrete time domain, but the model they proposed only has a dynamic intercept terms, which may be not sufficient enough for our purpose.

In this chapter, we will review basis of Bayesian dynamic models and how they can be used to examine the effect of ENSO on DSFW. We will explore different dynamic structures, extend and modify the models to the data at hand. To obtain a fast inference in large amount of data, INLA is performed. Model fitting criterion such as Deviance Information Criterion (DIC) is also computed.

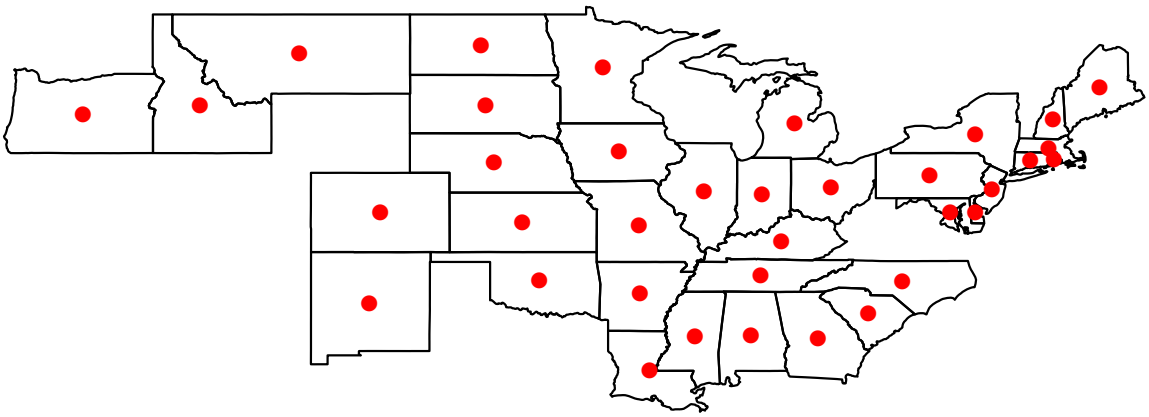

Figure 2.1: *The Map of 36 states in U.S. which has DSFW data available. The red dots show the centroid points of each states.*

2.2 Exploratory Analysis and Literature Reviews

2.2.1 Exploring the Data

The DSFW data is collected over 18 years (1996-2013) in United States (https://quickstats.nass.usda.gov/). An averaged DSFW for all sampler regions are obtained by state government, but not all states have DSFW data available in public. Only 36 states has sufficient data and the location of those states are shown in Figure 2.1. For example, State of California only have 4 years of data until 2013. Then, the State of California is truncated from our dataset.

More than 50 years record of El Niño Southern Oscillation is available for our study. Multiple measurements are taking to evaluate ENSO. One of the commonly used index, the Oceanic Niño Index (ONI) measures the difference in sea surface temperature in the Niño 3.4 region, which is considered to be most ENSO-representative (Bamston et al. 1997). For our research purpose, we only use ONI from the 1996 to 2013, which matches

the period of available DSFW data.

Noting that DSFW is a weekly data but ENSO is a monthly data, we aggregated weekly DSFW to monthly DSFW in order to match the same time scale. Furthermore a percentage of monthly DSFW is used to account for the fact that each month has different number of weeks and days. The following method is used to obtain our response variable. First, DSFW is converted from original scale $[0, 7]$ to percentage in [0,1] by dividing 7. Then, as long as the first day of a week belongs to a month, the week is assigned to this month in aggregation. At last, the mean of all weekly percentage of DSFW within the same month is computed as monthly percentage DSFW. In addition, the DSFW is mainly collected from the planting season to harvest season. There is very little data during January, February, March and December in all states. Therefore, those four months of data is removed from dataset. Hence, only 144 months present in 18 years of data.

Besides the aggregation, a simple imputation is implemented for the missing values of the data. The missing value is imputed using the average between the year before and the after. If missing value show up consecutively in the data, the average between the nearest neighbors is used for imputation.

Prior to the analysis of our data, a check of distribution of our data should be conducted. Since the DSFW data is all converted to the percentage, the biggest concerned is its normality. Figure 2.2 indicates the distribution of Percentage DSFW is left skewed. Also, by the property of percentage, all data values are bounded by 0 and 1. The assumption of normality seems to be invalid in this case. A transformation should be applied. Baum (2008) demonstrates the logit transformation is one of the best way to deal with percentage data. This transformation is applied to DSFW data. Figure 2.3 shows the transformed DSFW data is much more bell shaped than percentage DSFW. The new

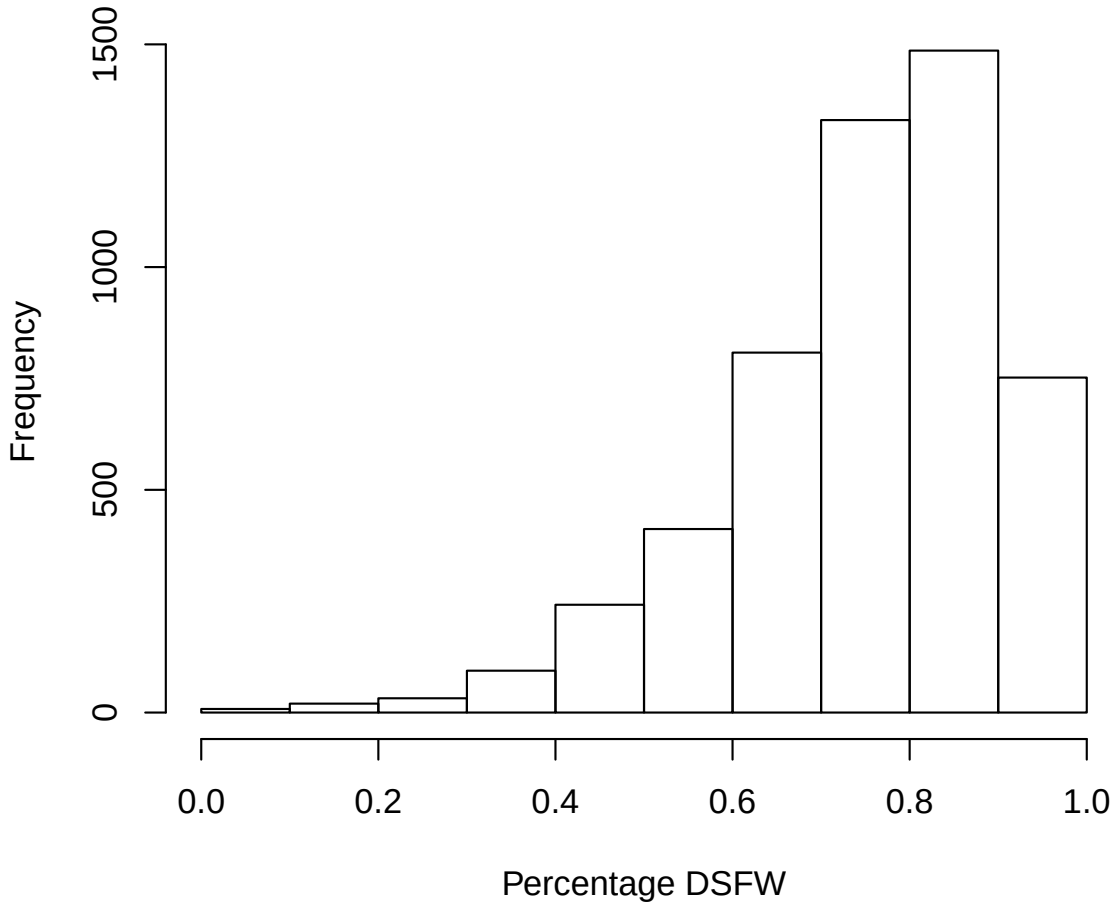

Figure 2.2: *A Histogram of Percentage DSFW*

response will be used for the further analysis in this research.

After pre-processing of the data, an initial linear regression model with transformed DSFW as response and ENSO as regressor is conducted. The time series of residual at two randomly picked states, Oregon, Kansas are shown on the Figure 2.4. The significance of ENSO is found in linear model with p-value= 0.00004 assuming the OLS assumption is

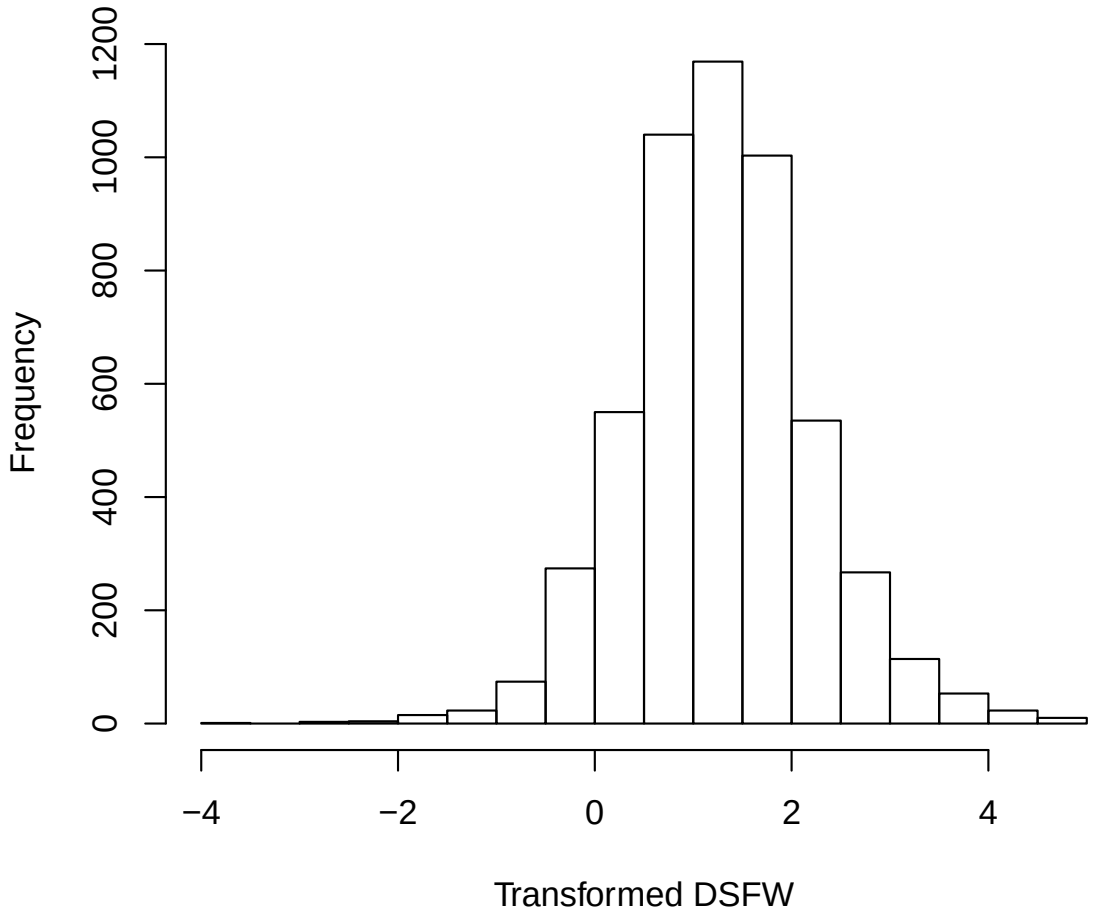

Figure 2.3: *A Histogram of Transformed DSFW*

met. However, in fact residuals admit an autocorrelation and seasonality as shown ACF and PACF plot of Figure 2.5. The Ljung-Box-Pierce tests for each states at lag 1 receive p-value= 0.000216 and 1.5×10^{-9}. It is a clear indication of presence of autocorrelation given ENSO present in the model. Periodogram of Kansas is plotted in Figure(2.6) and periodicity of around 0.125 is shown. It makes sense because there are 8 months per year

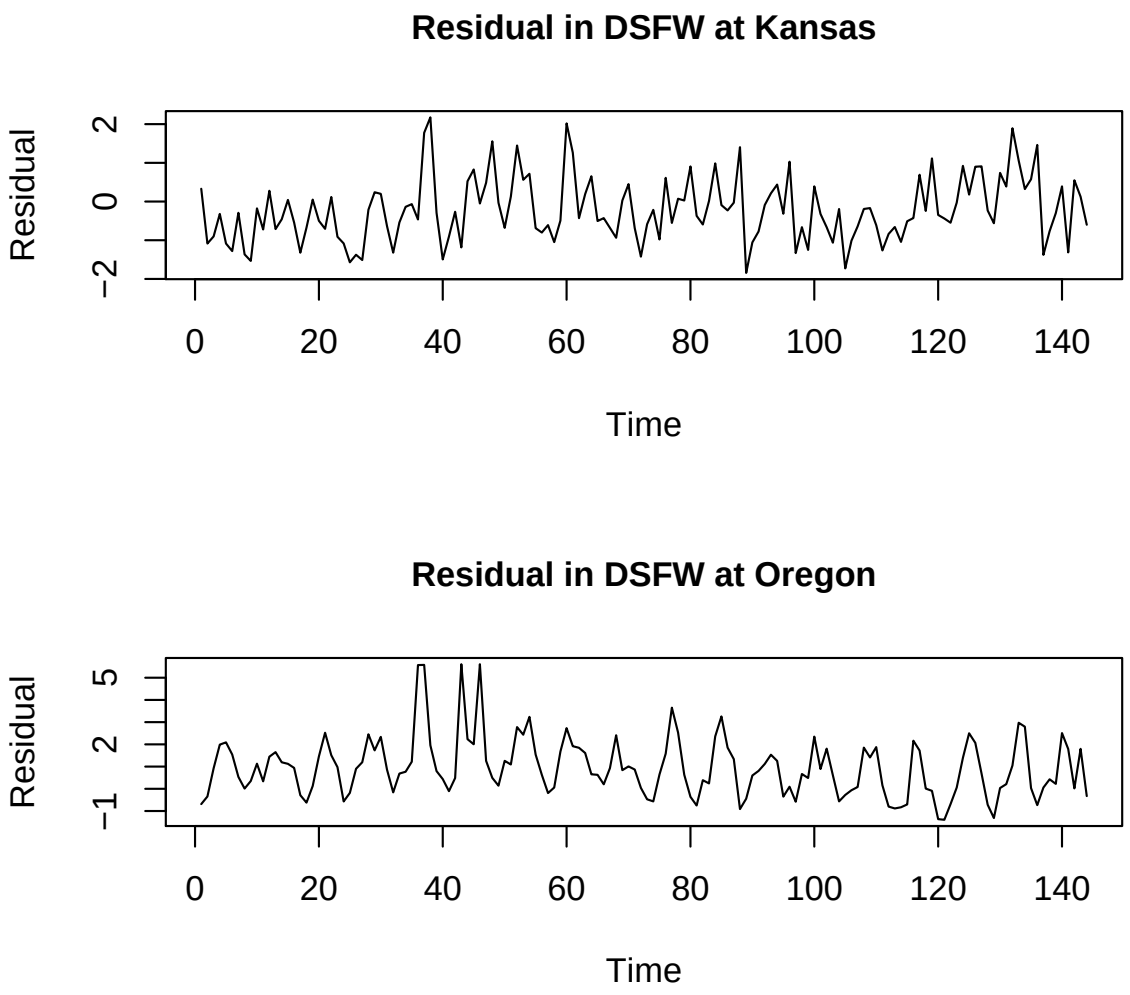

Figure 2.4: *Time series for residual DSFW at Kansas and Oregon*

after removing January, February, March and December.

To assess presence of spatial correlation structure in residual, a example of four randomly picked time points, October 1997, May 2000, August 2006 and July 2013 is given and a cross-sectional spatial plot is presented in Figure 2.7 at all four time point. Moran's I tests of spatial randomness are conducted with resulting p-values

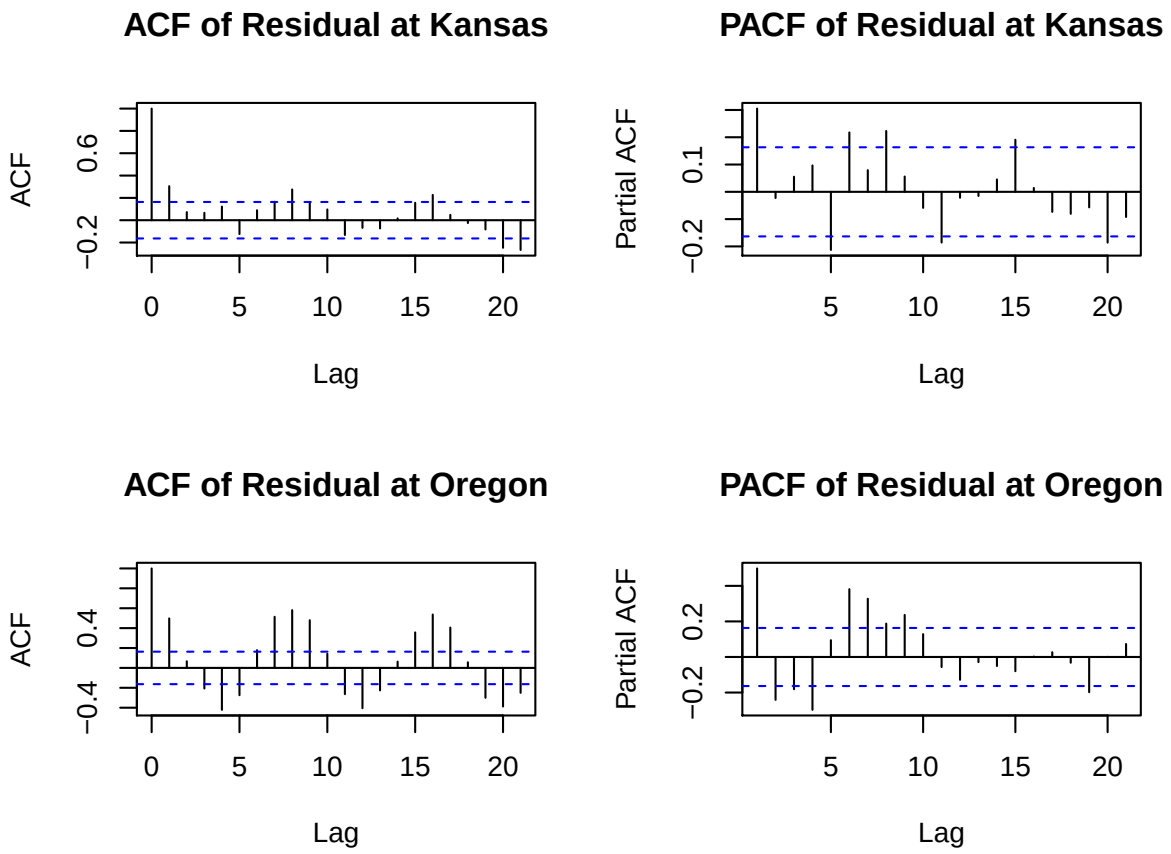

Figure 2.5: *ACF and PACF for residual at Kansas and Oregon*

$= 2.56 \times 10^{-5}, 1.6 \times 10^{-3}, 1.8 \times 10^{-4}$ and 1.3×10^{-7}. The significance of Moran I test confirms the need to incorporate spatial autocorrelation among states in the regression modelling. Figure 2.7 also demonstrates various patterns of clustering at different time point. It indicates spatial structure may interact with time.

2.2.2 Bayesian Dynamic Models

Bayesian dynamic model as described in the introduction of this chapter is also known as state-space model (Migon et al. 2005), which can be seen as a generalization of regression

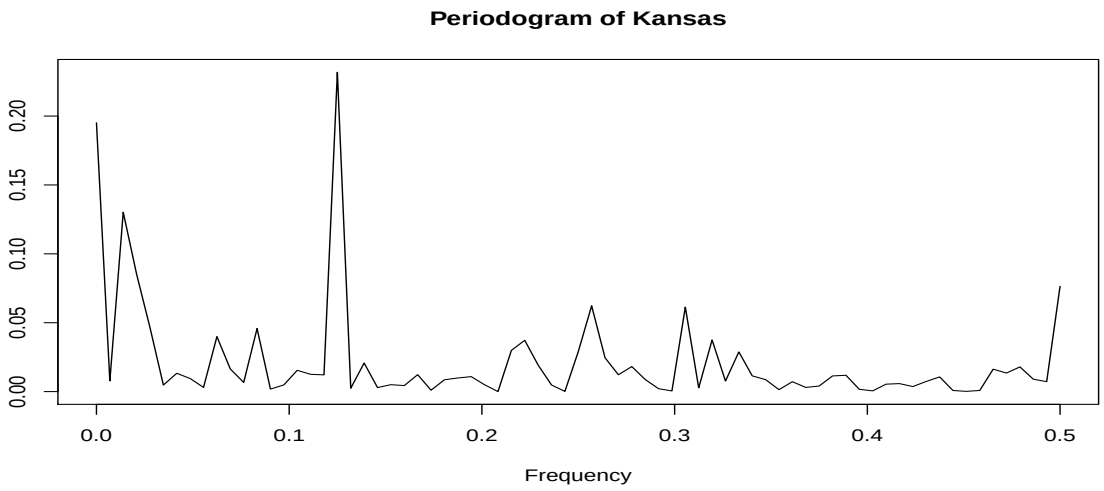

Figure 2.6: *Periodogram for residual DSFW in Kansas*

model. It consists of two parts, the observation equation and the state equation. By allowing a time varying and space-time varying coefficient, the state equation governs the temporal evolution of regression coefficients.

In the linear Gaussian case, the observation equation can be generally described by equation 2.1. The equation can be seen to describe the relationship between the response variable and predictors and underlying spatio-temporal processes. Noting that when a state-space model is built for spatio-temporal data, the response variable $\mathbf{y}_t$ is a vector of dimension n where the number of locations is n and the number of time points is k. Furthermore, the predictor $\mathbf{x}$ is also a vector, whose dimension depends on how many predictors present in the model. The design matrix $\mathbf{F}$ is corresponding to the dimension of $\mathbf{x}$ at each time point.

$$\mathbf{y}_t = \mathbf{F}_t{}'\mathbf{x}_t + \boldsymbol{\nu}_t,\ \boldsymbol{\nu}_t \sim N(\mathbf{0}, \mathbf{V}_t),\ t = 1, \ldots, k \tag{2.1}$$

The State equation can be generally described by equation 2.2. It can be seen to

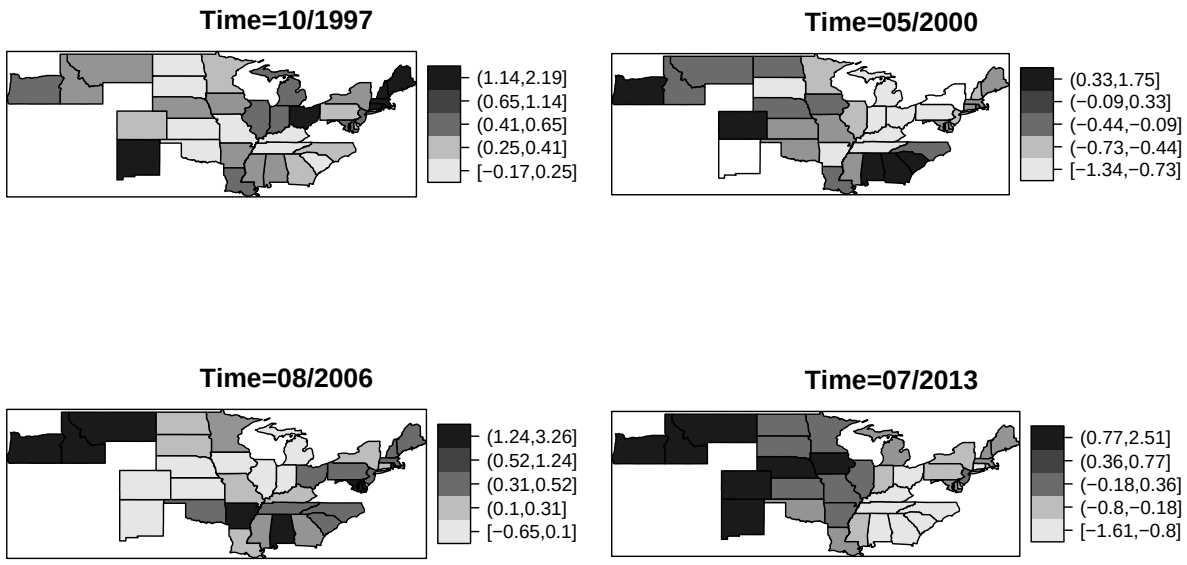

Figure 2.7: *DSFW data of 36 states in 4 time points*

describe the underlying processes which govern the observed process. Noting that the observed process can be governed by multiple processes, so that the there can be multiple state equations. For each state equation, the current state vector $\mathbf{x}_t$ is generally depends on the previous states. It introduces a relationship between the parameters in the observed equation but confines this relationship to have a markov property. In many cases, the state equation is used to specify parameters which are varying at different space and time. It is not estimable if every space and time has their own coefficients. The state equation serves as a restriction to allow estimation but still maintain flexibility of the model. On the other hand, the model is specified by $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$. People may let $\mathbf{V}_t = \mathbf{V}$ and $\mathbf{W}_t = \mathbf{W}$, which means the covariance for the innovation process does not change over time. This assumption is adopted by this research as well to reduce complexity of the

model.

$$\mathbf{x}_t = \mathbf{G}_t \mathbf{x}_{t-1} + \boldsymbol{\omega}_t,\ \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t),\ t = 2, \ldots, k \tag{2.2}$$

2.2.3 Gaussian Markov Random Field

In view of the fact that Days Suitable for Fieldwork (DSFW) are collected on an irregular lattice on space and evenly spaced lattice on time, we will first consider the data as a realization of a discrete spatio-temporal stochastic process. To properly study the interactive relations of multiple spatio-temporal processes, a complex hierarchical Bayesian model will be applied and Gaussian Markov Random Field plays a central role in this framework (Rue and Held, 2005).

Any finite dimensional subset of a Gaussian Random Field $x_s, s \in D$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$, is an n-dimensional random vector following a multivariate normal distribution, i.e. $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Markov property establishes that the fully conditional distribution of a random variable x_i depends only on its neighborhood values, i.e.

$$p(x_i | \{x_j : j \neq i\}) = p(x_i | \{x_j : j \in \boldsymbol{N}_i\})$$

where $\boldsymbol{N}_i$ is neighborhood of variable x_i. Conditional independence is an intermediate step between independent and full dependent structure. Gaussian Markov Random Field is Gaussian Random Field with Markov property and the local specification can determine the joint distribution (Rue and Held, 2005). One of the simplest GMRF is the conditional autoregressive of order 1 process in one dimensional space. It is often expressed as

$$x_t = \phi x_{t-1} + \epsilon_t,\ \epsilon_t \overset{iid}{\sim} N(0, 1),\ |\phi| < 1, \tag{2.3}$$

where t represents time. The assumption of conditional independence leads to a specific full conditional form given as

$$x_t|x_1, ..., x_{t-1} \sim N(\phi x_{t-1}, 1),\ t = 2, \ldots, n. \tag{2.4}$$

The Markov property also produces a sparse precision matrix (inverted variance-covariance matrix). Theorem 2.2 in Rue and Held (2005) provides a explicit link between conditional independence and zero values in the precision matrix, i.e. precision matrix $\mathbf{Q}_{ij} = 0$ if and only if x_i and x_j are conditional independent upon $\mathbf{x}_{-ij}$. In spatial or temporal case, the relationship between locations or time points usually is encoded to a graph, let's say $\mathcal{G}$. This result indicates the nonzero pattern in precision $\mathbf{Q}$ can be determined by the graph $\mathcal{G}$ and it is therefore convenient to specify the likelihood of a GMRF $\mathbf{x}$ in the form of precision matrix as following:

$$\pi(\mathbf{x}) = (2\pi)^{-n/2}|\mathbf{Q}|^{1/2} exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{Q}(\mathbf{x}-\boldsymbol{\mu})\right\}. \tag{2.5}$$

Usually only $\mathcal{O}(n)$ of n^2 entries of $\mathbf{Q}$ are non-zero, so $\mathbf{Q}$ is sparse (Rue et al., 2009). Cholesky decomposition can be used to factorize $\mathbf{Q}$ as $\mathbf{L}\mathbf{L}^T$, where $\mathbf{L}$ is lower triangle matrix. Thanks to the global Markov property (Theorem 2.8 in Rue and Held 2005), the sparseness of precision $\mathbf{Q}$ is inherited to $\mathbf{L}$. One of the primary computational cost in sampling in the MCMC or estimation in Maximum likelihood approach is due to the inverting variance-covariance matrices. While the approach for GMRF mostly directly works on precision $\mathbf{Q}$ modeling and can somewhat mitigate computation intensity by avoiding matrix inverse calculation.

The properties of GMRF not only provide a link between hierarchical models and sparse matrix algorithm, but also indicate a significant computational reduction from full

dependent structures. GMRF can be naturally implemented to MCMC algorithm, however, for large spatial or spatio-temporal data sets, MCMC exhibits poor performance in term of effective sample size. In order to achieve a high effective sample size, a much longer MCMC samplers may be required, which lead to an expensive computational cost. Integrated Nested Laplace Approximation (INLA) is a fast alternative to MCMC. It reduces computational time in magnitude and therefore provides a viable Bayesian inference.

2.3 INLA Based Bayesian Framework with Complete Proof of Key Approximation

As it's mentioned in the previous section, ENSO vs DSFW data contains 36 locations in 144 time points. Applying a Bayesian dynamic models using MCMC can be computationally intensive. Instead, we choose to use approximated Bayesian inference.

Integrated Nested Laplace Approximation (INLA) is an approximate Bayesian inference mostly for a subclass of structured additive model, named Latent Gaussian model. Under this framework, suppose the distribution of response belongs to an exponential family and the mean μ_i of the response distribution is linked to predictors in an additive structure. Denote link function as $g(\cdot)$, then

$$g(\mu_i) = \beta_0 + \sum_{m=1}^{M} \beta_m u_{mi} + \sum_{i=1}^{L} f_l(z_{li}), \tag{2.6}$$

where β_0 is the intercept; $\beta = (\beta_1, \ldots, \beta_M)$ is the linear effect of covariates $\mathbf{u} = (u_1, \ldots, u_M)$ and $f = (f_1(\cdot), \ldots, f_L(\cdot))$ is a collection of functions defined in term of a set of covariates $\mathbf{z} = (z_1, \ldots, z_L)$. The term $f(\cdot)$ can be considered as a smooth function, a nonlinear effect of covariates, time trends, seasonal effect, a random effect or a spatial random effect, etc.

(Blangiardo and Cameletti, 2016). Latent Gaussian model can accommodate wide variety of models, such as regression (generalized regression) models, dynamic models and spatial and spatial-temporal models are all Latent Gaussian model. Note that the functional term $f(\cdot)$ increase flexibility for the form of covariates.

The latent field x contains all latent Gaussian variables, say $\{f(\cdot)\}$, $\{\beta\}$ and $\{\epsilon\}$. Due to Gaussian distribution and discrete indexed nature of x, x is often structured to be a GMRF. All properties of GMRF therefore applies to latent field x and the parameters in marginal distribution of x, for example θ in precision $\mathbf{Q}$, are hyper-parameters. A vector of all hyper-parameters denoted as $\boldsymbol{\theta}$ do not need to have a Gaussian prior. The observed process y, latent field x and hyper-parameters $\boldsymbol{\theta}$ with their associate distributions conditionally specified constitute a Bayesian hierarchical model.

For the simplicity, we assume our response or a transformed response follows a normal distribution. Hence, let's define a Bayesian hierarchical model as three stages,

The data model : $y_i|\mathbf{x}, \boldsymbol{\theta} \overset{iid}{\sim} N(\beta_0 + \sum_{m=1}^{M} \beta_m u_{mi} + \sum_{i=1}^{L} f_l(z_{li}), \sigma_\epsilon^2 \mathbf{I})$,

The latent field: $\mathbf{x}|\boldsymbol{\theta} \sim N(\mu_{\mathbf{x}}(\boldsymbol{\theta}), \mathbf{Q}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}))$,

The hyper prior: $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

Note that all parameters such as $\beta_0, \beta_m, f_l(\cdot)$ are included in $\mathbf{x}$ for each data point and all hyper-parameters such as σ_ϵ^2 and parameters used to define f_l are all included in $\boldsymbol{\theta}$. Each observation $y_i, i \in I$ is conditionally independent given $\mathbf{x}$ and $\boldsymbol{\theta}$. The latent field can have a very large dimension. Since it is assumed to be a GMRF with a sparse precision matrix $\mathbf{Q}_x(\boldsymbol{\theta})$, the computation reduction is achieved. The hyper-prior is pre-determined and usually a vague prior.

The common objective of Bayesian inference are marginal posterior distributions for each element of latent field x and for each element in the hyper-parameter $\boldsymbol{\theta}$. That is,

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \tag{2.7}$$

and

$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{-j}. \tag{2.8}$$

One of the key steps of INLA is to construct nested approximations Rue et al. (2009),

$$\tilde{\pi}(x_i|\mathbf{y}) = \int \tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \tag{2.9}$$

and

$$\tilde{\pi}(\theta_i|\mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{-j}. \tag{2.10}$$

Here, $\tilde{\pi}(\cdot|\cdot)$ is an approximate conditional density of its arguments. First, INLA is going to approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$. The following proposition explains the form of this approximation.

Proposition 1.

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) = C\frac{\pi(x^*(\boldsymbol{\theta}), \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(x^*(\boldsymbol{\theta})|\boldsymbol{\theta}, \mathbf{y})} \tag{2.11}$$

Where $C^{-1} = \int\int \pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})d\mathbf{x}d\boldsymbol{\theta}$, $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ *is the Gaussian approximation of* $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. $\mathbf{x}^*(\boldsymbol{\theta})$ *is the mode of the conditional for* $\mathbf{x}$ *for a given* θ.

In fact, Rue et al. (2009) used the Proposition 1 form of (2.11), roughly mentioning the connection with Tierney and B. Kadane (1986). But there is no explicit proof of this relationship in proposition 1 above, to the best of our knowledge. Hence, we intend to fill this blank with detailed proof given in the appendix.

Second, $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ is computed by Laplace approximation again.

$$\begin{aligned}
\tilde{\pi}(\mathbf{x}_i|\boldsymbol{\theta}, \mathbf{y}) &= \frac{\pi((x_i, \mathbf{x}_{-i})|\boldsymbol{\theta}, \mathbf{y})}{\pi_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \\
&= \frac{\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})}{\pi(\boldsymbol{\theta}|\mathbf{y})} \frac{1}{\pi_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \\
&\propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})}{\pi_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \\
&\approx \frac{\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})}{\pi_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})}
\end{aligned}$$

where $\pi_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation to full conditional of $\mathbf{x}_{-i}$ and $\mathbf{x}^*_{-i}(x_i, \boldsymbol{\theta})$ is its mode point. Usually, the random variable $\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y}$ is reasonably normal. However, it can be computational expensive since $\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$ has to be recomputed for each x_i. Some dimensional reduction technique is required to speed up computation. Also, a simplied Laplace approximation is a default in INLA algorithm instead of Laplace approximation. The simplified Laplace approximation is based on Taylor's series expansion of the Laplace transformation $\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$. The more details is shown in originial INLA paper at (Rue et al., 2009).

A critical aspect of INLA is to explore $\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$ and $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ in a "nonparametric" way. Once we get $\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$ and $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, the integral in equation (2.7) can be approximated by

$$\tilde{\pi}(x_i|\mathbf{y}) \approx \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y}) \times \tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y}) \times \Delta_k \tag{2.12}$$

for some relevant integration points $\boldsymbol{\theta}_k$ with a corresponding set of weights Δ_k. The choice of weights Δ_k is also given in (Rue et al., 2009). More details of some of the proof in Laplace Approximation behind the INLA algorithm is given the in appendix.

INLA outperforms all other MCMC alternatives in computation speed. Operationally

we use INLA library (http://www.r-inla.org/) of R-software (R Core Team, 2018) for the estimation and implementation of the Bayesian dynamics models we developed in the later section.

2.4 Proposed Dynamics Models in Discrete Spatial Domain

Starting from this section, a more delicate analysis of the DSFW will be conducted and the results from INLA estimation will be shown and compared. It has been determined that the response variable, DSFW data is collected and averaged across each state and each month and its distribution is normal after the transformation. From the aspect of agriculture economics, we need to think two questions. One is if the spatial process of DSFW is a discrete or a continuous process. Another is if the impact of ENSO has a lag effect on the DSFW and how long is the lag.

One can argue that it is reasonable to describe the spatial domain of DSFW in an irregular lattice, which is simply the geographical map of each states. Others can argue that DSFW continuously distributed over the entire United States, as long as there is a farm. However, the way we are sampling the data, which is the data collection agency aggregated the DSFW data to each state in space, causes a discrete space-time model to be more make sense and more interpretable. In this section, we will work on our dynamics models in discrete spatial domain.

2.4.1 Model Without Time Varying Coefficients

In model 1, we are going to assume the coefficient of ENSO over DSFW does not change over time and space. i.e. Denote

$$\mathbf{y}_t = \Upsilon + \mathbf{z}_t\beta + \boldsymbol{\xi}_t + \ \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}) \tag{2.13}$$

$$\boldsymbol{\xi}_t = \alpha\boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_t, \ \boldsymbol{\omega}_t \sim PGMRF(\mathbf{0}, \mathbf{W}^{-1}) \tag{2.14}$$

Where $\mathbf{y}_t = (y_{t1}, \ldots, y_{tS})'$ is a vector the transformed DSFW at time $t = 1, \ldots, 144$, $s = 1, \ldots, 36$; z_t is the ENSO Oceanic Niño Index on time t; Υ is the intercept of the model and β is the main effect of DSFW; $\boldsymbol{\epsilon}_t$ is the Gaussian measurement error term since a Gaussian distribution is assumed for response. $\boldsymbol{\xi}_t$ describes the underlying spatio-temporal process. This process is governed by the state equation 2.14. At each time t, the error vector $\boldsymbol{\omega}_t = (\omega_{t1}, \ldots, \omega_{tS})'$ follows a Proper Gaussian Markov Random Fields(PGMRF) (See e.g. Vivar and Ferreira 2009). Matrices $\mathbf{W}$ describes the spatial covariance structure of the state at time t. Precision matrices $\mathbf{W^{-1}}$ are modeled as $W^{-1} = \tau(\mathbf{I} - \frac{\phi}{\lambda_{\mathbf{max}}}\mathbf{C})$ and $\mathbf{C}$ is a structure matrix defined as

$$C_{k,l} = \begin{cases} c_k, \ \text{if } k = l, \\ -h_{k,l}, \ \text{if } k \in d_l, \\ 0, \ \text{otherwise}, \end{cases}$$

where d_l is the set of neighbors of area $l, h_{l,k} > 0$ is a measure of similarity between areas k and l, if area k and l are adjacent then $h_{l,k} = 1$; $c_k = \sum_{l \in d_k} h_{k,l}$; λ_{max} is maximum eigenvalue of matrix $\mathbf{C}$; τ is a scale parameter and $0 \leq \phi < 1$ controls the degree of spatial correlation.

The observation equation in model 1 assumes the ENSO and DSFW has a linear relationship. The state equation is a typical Autoregressive of order 1 process (AR1). AR1 process is very commonly used for creating a stationary dependent structure which is only conditionally depends on 1 previous state. This type of structure fall between an independent assumption and fully dependent structure. It gains a lot of computational advantage but also has certain degree of flexibility. This is the reason why many practitioners use this process and we choose this process to begin with.

Besides, there is a spatial pattern within the state vector which is described in $\mathbf{W}$. The initial state is assumed to be $\boldsymbol{\xi}_1 \sim N(\mathbf{0}, (\tau(1-\alpha^2)))^{-1}$. It is a similar set-up with Cameletti et al. 2013's equation (1) and (2). The differences is they used Stochastic Partial Differential Equation approach to construct a continuous spacial domain. The form of the posterior distribution is therefore given in equation (8) in Cameletti et al. 2013.

Fitting this model using INLA, we get a posterior mean for the effect of ENSO (β) as -0.077 and 95% credible interval (-0.108, -0.046). The credible interval of the effect of ENSO does not include 0. This indicates seems like the effect of ENSO is significant at 5% level. The DIC for this model is 13432.77. The model suggests there is a negative relationship between ENSO and DSFW, means that when then ENSO ONI is decreasing, the DSFW would increase on average. In the line with the previous discussion, the relationship between DSFW may be more complex than a simple linear relation. The intercept of the observation equation is affect by a "fix" part, the β_0 and a "random" part, the $\boldsymbol{\xi}_t$. The dynamic of the model comes from the time varying coefficients $\boldsymbol{\xi}_t$ which can be considered as a random intercept. The random intercept is only able to reflex the dynamic nature of DSFW itself, not the relationship between DSFW and ENSO.

2.4.2 Model With Time Varying Coefficients

In the next, model 2 will add a time-varying coefficient to describe the dynamic relationship between ENSO and DSFW. To do so, another state equation is added to the model as follow,

$$\mathbf{y}_t = \Upsilon + \mathbf{z}_t\beta_t + \boldsymbol{\xi}_t + \ \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}) \tag{2.15}$$

$$\boldsymbol{\xi}_t = \alpha_1\boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_t, \ \boldsymbol{\omega}_t \sim PGMRF(\mathbf{0}, \mathbf{W}^{-1}) \tag{2.16}$$

$$\beta_t = \beta_0 + \alpha_2\beta_{t-1} + \epsilon_{2t}, \ \epsilon_{2t} \sim N(0, \sigma_{\epsilon_2}^2) \tag{2.17}$$

The observation equation 2.15 is almost the same with observation equation 2.13 besides the coefficient β is able to change over time. The state equation 2.16 is the same with state equation 2.14 in the previous model. The second state equation 2.17 is to regulate the time-varying coefficient β_t to form a (AR1) patten. However, the difference between the state equation 2.16 and 2.17 is the intercept. Since state equation 2.16 describes underlying spatio-temporal random effect in the observation equation, the expectation of the process should be 0. However, the coefficient of the ENSO on DSFW may be away from 0. This is exactly what we want to test for this study. Hence, an intercept β_0 is add to the state equation 2.17.

Fitting model 2 using INLA, the DIC of the model 2 decrease to 8999.002, which indicates the fit is much better than model 1. The posterior mean for the main effect β_0 of ENSO is 0.214 with a 95% credible interval $[0.126, 0.304]$, which means the effect of ENSO is significant over DSFW. Figure 2.8 demonstrates the posterior mean of coefficients β_t of ENSO over time and compare with the ENSO itself. It seems that the oscillation of the coefficients for ENSO is very steep and potentially it has a lag between the peak of ENSO and peak of its coefficients.

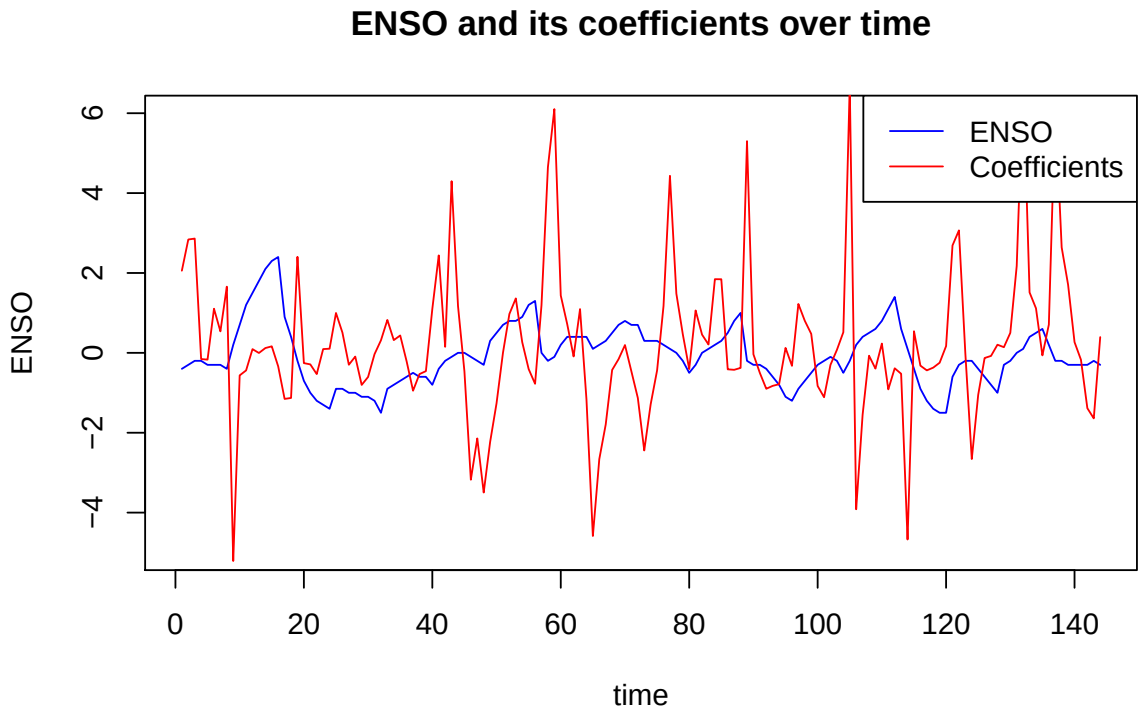

Figure 2.8: *The ENSO and the posterior mean of its coefficients over time*

2.4.3 Model With Space and Time Varying Coefficients

Fellowing the thoughts in the model 2, it is natural to think the effect of ENSO may have a spatio-temporal pattern. Since the El Niño and La Niña happened over pacific ocean, it is reasonable to assume the effect of ENSO may be different from west coast to east coast. However, the problem is that ENSO is an index for each month over the globe. As a result, the ENSO isn't measured over different locations. If we want to evaluate ENSO's impact on different space, an assumption must be made. We are going to assume the effect of ENSO is connected the same way with the DSFW itself. Denote model 3 as following,

$$\mathbf{y}_t = \Upsilon + \mathbf{Z}_t\boldsymbol{\beta}_t + \boldsymbol{\xi}_t + \ \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}), \tag{2.18}$$

$$\boldsymbol{\xi}_t = \alpha_1\boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_t, \ \boldsymbol{\omega}_t \sim PGMRF(\mathbf{0}, \mathbf{W}^{-1}), \tag{2.19}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_0 + \alpha_2\boldsymbol{\beta}_{t-1} + \ \boldsymbol{\epsilon}_{2t}, \quad \boldsymbol{\epsilon}_{2t} \sim PGMRF(0, \mathbf{W}^{-1}), \tag{2.20}$$

where the observed process y_t is assumed to be governed by two processes, the intercept process has a state equation 2.19 the same with state equation 2.14 and 2.16 in model 1 and model 2; $\mathbf{Z_t}$ is a matrix with the z_t value on the diagonal, otherwise 0; The effect $\boldsymbol{\beta}_t = (\beta_{t1}, \ldots, \beta_{tS})'$ becomes a vector which represents a vector of coefficients at each time point. Within a time point, each elements of the vector represents an coefficient for a single location. This process is governed by the state equation 2.20 with a spatial covariance structure $\mathbf{W}$ same with state equation 2.19. Since a new process is added to the effect $\boldsymbol{\beta}$ of the model, the joint posterior density has to be computed. The details of the joint posterior is included in the appendix.

Fitting the model in INLA, the DIC of model 3 is 13402.2, which is in-between model 1 and 2's DIC. Three time points are randomly selected (34,64,91). Figure 2.9 demonstrates the posterior mean of effect over the space. Even though the spatial patterns seems to be different at each location, there is no obvious pattern inherited through time. In this method, there is an different parameter for the effect of ENSO at every ENSO at each time and space. One may argue that this is an over parameterized model and the increasing DIC provide evidence to support the statement. We will only demonstrate the results of this model and continue to explore in another way.

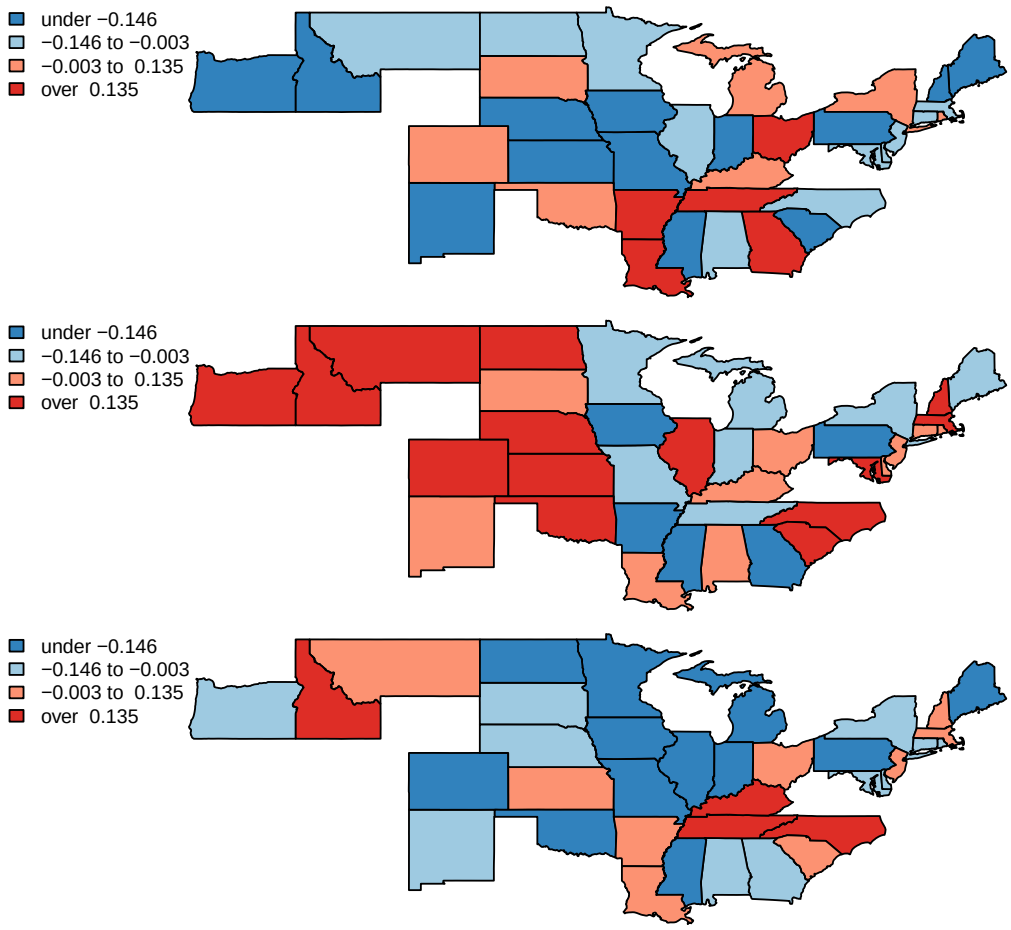

Figure 2.9: *The posterior mean of ENSO coefficients over space at the time point 34, 64 and 91*

2.5 Proposed Dynamics Models in Continuous Spatial Domain

For a practitioner in spatial/ spatio-temporal models, it is sometimes reasonable to assume a continuous spatial domain rather than discrete domain. When the predicted location is specified but not on the lattice or one want to has a more complexed spatio-temporal covariance structure, a model built on the continuous spatial domain is in need. In this

research, one can argue that DSFW is a continuous spatial process since there are farms almost everywhere over the United States. The sampled DSFW is an averaged results from many farms in a state and many states the USDA does not provide DSFW in a finer spatial resolution than the state level. It forces us to consider a lattice at the state level in the previous section. The purpose of this section is to retain a continuous spatial domain from the sampled DSFW data and then built a model on top of it to see if any improvement may be obtained through this process.

A problem in a continuous spatial domain usually requires a much bigger computational burden than a in a discrete spatial domain. To reduce the computational burden associated with GF, Lindgren et al. (2011) provides an explicit link between Gaussian Random Fields (GF) and Gaussian Markov Random Fields (GMRF) using stochastic partial differential equations (SPDE). In this approach, the GF is discretised by Delaunay triangulation and the new discrete random field is considered to be a GMRF. The SPDE technique redefines the Matérn GF as a finite linear combination of basis functions on each triangulated vertices. Hence, the spatial/spatio-temperal random effect is approximated by a basis $\mathbf{B}$, which is a sparse matrix with unit elements for matching triangle vertices and zero's elsewhere (Lindgren et al. 2011). The computation reduction happens on the zero pattern of precision matrix (Rue et al. 2009). GMRFs have very good computational properties, which is of major importance in Bayesian methods. This is further enhanced by the link to INLA. Cameletti et al. (2013) provides an example in continuous spatial domain and discrete time using SPDE approach in INLA, however, it only contains an intercept process and the effect of coefficients are considered not changing over time and space. In later part of this section, I have

In Cameletti et al. (2013), the spatial domain is subdividing into a set of non-intersecting triangles, where any two triangles meet in most one common edge or corner. Three cor-

ners of a triangle are named as vertices. The initial vertices are placed at the location where the data are collected or the location you want to predict. Additional vertices are added to satisfy the overall constraints of the triangles, such as maximally allowed edge length, and minimally allowed angles. This is the standard solution in engineering for solving partial differential equations using finite element method. In this research, the initial vertices are placed at the centroid point of each states, where is assumed to the location DSFW observed. However, we does not place additional triangles inside of the spatial range of United States. Since the location of each states forms a natural lattice, we does not want to artificially separate the natural neighborhood by adding additional vertices within the map.

Figure 2.10 demonstrates this Delaunay triangulation. The spatial domain contains 54 vertices. The red dots are the assumed locations for each DSFW observations, which are the centroid points of each states. To avoid a boundary effect, the triangulated region is a little bit bigger than the original spatial domain, so that additional vertices are added outside of the border.

Lindgren et al. (2011)'s link from GF to GMRF works well if the GF has a matérn covariance structure. Denote $\mathbf{X}(\mathbf{s}) = x(s), s \in D \subseteq \mathbb{R}^2$ denotes a GF with a matérn covariance function. In particular, the precision matrix Q_s of the GMRF is defined by Equation (10) of Lindgren et al. (2011) as a function of κ^2 and smoothness $0 < \alpha \leq 2$. The $\nu = \alpha - 1$ is the smoothness parameter commonly used for the matérn covariance function. The SPDE approach essentially uses a finite element representation to define this field as a linear combination of basis functions defined on a triangulation of the domain D (Cameletti et al. 2013). Given the triangulation, the basis function representation is given by

$$\mathbf{X}(\mathbf{s}) = \sum_{l=1}^{n} \psi_l(\mathbf{s})\omega_l, \tag{2.21}$$

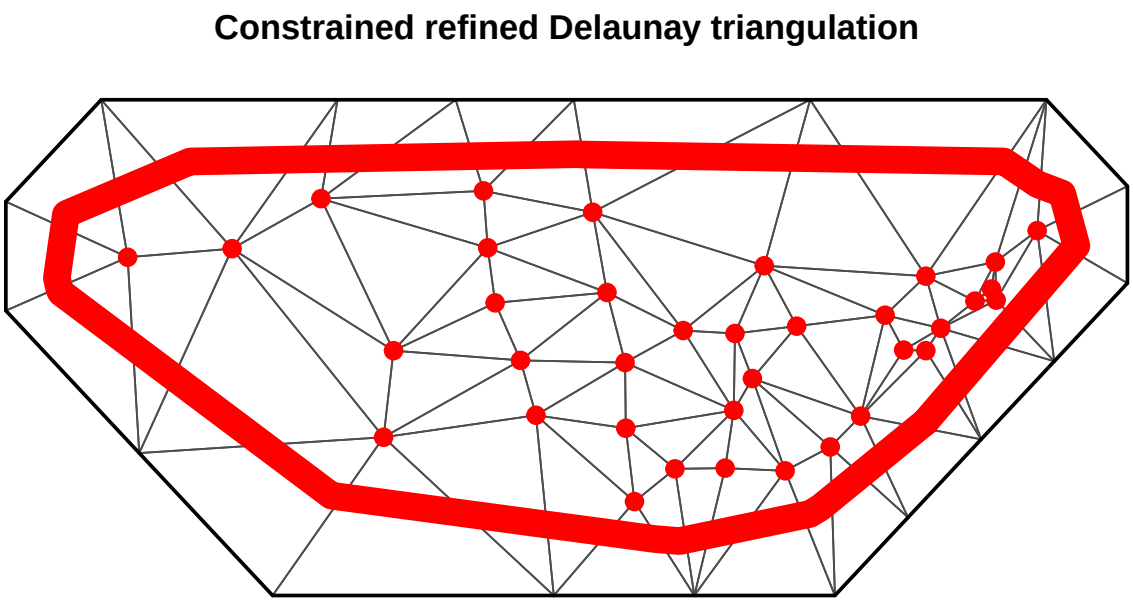

Figure 2.10: *Delaunay triangulation for the spatial field of DSFW. Centroid points of each states are red dots and a convex hull of the border is draw in red line.*

where n is total number of vertices, $\psi_l(\mathbf{s})$ are the basis functions and ω_l are Gaussian distributed weights. The functions $\psi_l(\mathbf{s})$ are chosen to be piecewise linear on each triangle, i.e. $\psi_l(\mathbf{s})$ is 1 at vertex l and 0 at all other vertices.

Hence, we assumes DSFW follows a GF with matérn covariance function and but the smoothness parameter has to be set first. In this research, multiple smoothness parameter is tried out and it was found that it does not influence the resulting model and the conclusion too much. For simplification, we set the smoothness parameter of all SPDE model to be 1.5, which is corresponding to a matérn field with 0.5 smoothness.

2.5.1 Extending the First Model

Collecting all the observations measured at time t in a vector $\mathbf{y}_t = (y_{t1}, \ldots, y_{tS})'$, we write model 4 as following,

$$\mathbf{y}_t = \Upsilon + z_t\beta + \boldsymbol{\xi}_t + \ \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}) \tag{2.22}$$

$$\boldsymbol{\xi}_t = \alpha\boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_t, \ \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{Q}_s^{-1}) \tag{2.23}$$

The set-up of the model 4 is very similar with model 1 except that the assumption that $\boldsymbol{\xi}_t$ is spatial-temporal process in a continuous spatial domain and discrete time domain. This is also exactly the same model with equation (1) and (2) in Cameletti et al. (2013). The covariance structure is not a PGMRF but approximated by the SPDE approach. We also need to assume the covariance structure stays the same over time, which is the same assumption previously presented in the discrete time.

Fitting model 4 in INLA, the DIC values is 11950. The posterior mean for the β is -0.068 and 95% creditable interval is (-0.267, 0.130). The creditable interval contains zero, which indicates a non-significant effect of ENSO over DSFW. This result contradicts with the result we have in model 1-3. However, as the DIC increases significantly from the best previous model (model 2), we continue to extend our model to have a better fit.

2.5.2 Extending Model with Time Varying Coefficients in Continuous Domain

The first extension from model 1 is a used model in Cameletti et al. (2013). Our contribution in this section is to extending model 4 and apply a time-varying coefficients for ENSO using INLA. Similarly to model 2, the coefficient β is able to change over time.

The state equation 2.25 is the same with state equation 2.23 in model 4. The second state equation 2.26 is to regulate the time-varying coefficient β_t to form an AR1 pattern. An intercept β_0 for the β_t process is added to the state equation to allow the process mean deviate from 0.

Hence, model 5 can be written by

$$\mathbf{y}_t = \Upsilon + z_t\beta_t + \boldsymbol{\xi}_t + \ \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}) \tag{2.24}$$

$$\boldsymbol{\xi}_t = \alpha_1\boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_t, \ \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{Q}_s^{-1}) \tag{2.25}$$

$$\beta_t = \beta_0 + \alpha_2\beta_{t-1} + \epsilon_{2t}, \ \epsilon_{2t} \sim N(0, \sigma_{\epsilon_2}^2) \tag{2.26}$$

Model 5 corresponds to the model 2 in the previous section. Fitting model 5 in INLA, the DIC value is 11937.67. The main effect β_0 has posterior mean of 0.067 and 95% creditable interval (-0.204, 0.337), which is not a significant effect over DSFW. Figure 2.11 demonstrates the posterior mean of ENSO's coefficients β_t over time. The ENSO's time varying effect shown in figure 2.11 seems to be less than the one in model 2, comparing with figure 2.8. The DIC value of model 5 is slightly better than model 4, but still much bigger than model 2.

2.5.3 Extending Model with Space and Time Varying Coefficients in Continuous Domain

Continuing the line of thoughts and for the same reason mentioned in model 3's work, we can extend the effect with a coefficient changing over time and space. Now, the effect of the ENSO(i.e. $\boldsymbol{\beta}$) is assumed to have a continuous spatial process in every time point. We are also going to assume this spatial domain is connected the same way with DSFW

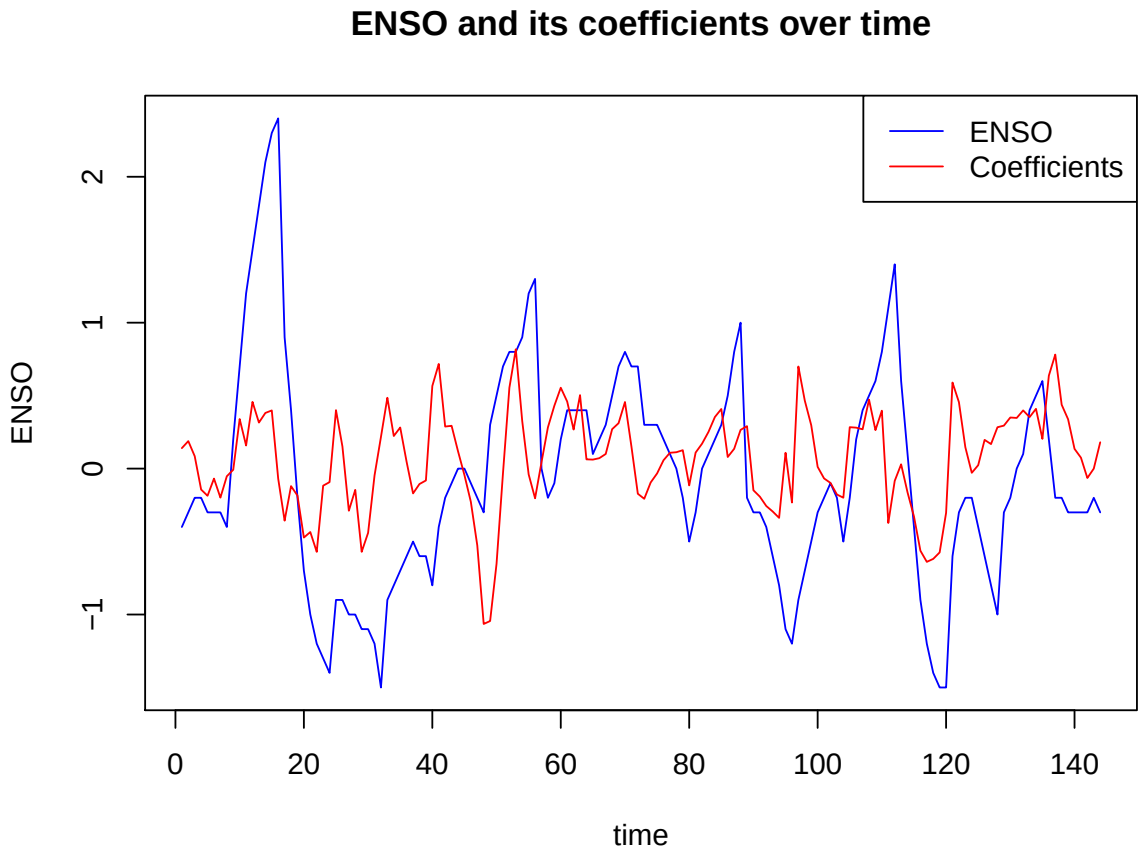

Figure 2.11: *Under SPDE model, the ENSO and the posterior mean of its coefficients over time*

itself. From temporal evolution side, the same AR1 structure is used with state equation 2.29.

Thus, model 6 can be written as

$$\mathbf{y}_t = \Upsilon + \mathbf{Z}_t\boldsymbol{\beta}_t + \boldsymbol{\xi}_t + \ \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}), \tag{2.27}$$

$$\boldsymbol{\xi}_t = \alpha_1\boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_t, \ \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{Q}_s^{-1}), \tag{2.28}$$

$$\boldsymbol{\beta}_t = \beta_0 + \alpha_2 \boldsymbol{\beta}_{t-1} + \ \boldsymbol{\epsilon}_{2t}, \quad \boldsymbol{\epsilon}_{2t} \sim N(0, \mathbf{Q}_s^{-1}), \tag{2.29}$$

Model 6 corresponds to model 3 in the previous section, except the spatial covariance structure W is replaced by a ${Q_s}^{-1}$. The new covariance structure is also approximated by the SPDE approach introduced earlier. Fitting the model 6 in INLA, the DIC is 11862.56. The posterior mean of the ENSO's main effect is -0.073 with 95% creditable interval (-0.343, 0.196). Figure 2.12 presents the posterior mean of ENSO's effect at time point 34, 64 and 91. To compare with the result in the discrete spatial domain, we select the same time with figure 2.9. The result in figure 2.12 seems to be very different with figure 2.9 and also it seems to have certain level of randomness among different time.

In this section, three models are constructed to try to improve the results from the previous section and provide means for prediction in any point in space. However, the results by considering a continuous spatial domain is not as good as considering a discrete spatial domain if we use DIC as model selection criterion. Instead of giving ENSO a significant effect over DSFW in model 1-3, model 4-6 suggests otherwise. One can also argue that the DIC among model 4-6 is significantly higher than the best model (model 2), so that the result is not as trustworthy as model 2. At this point, the main reason to develop a continuous spatial model for this type of problem is the prediction. Given the predicted location, we only need to add those location as vertices in the Delaunay triangulation and then fit the model again. Then, the prediction can be easily computed. We will not demonstrate this in this research.

2.6 Extending Model with Lag Effect

From the previous section, we determine the model 2, which is a time-varying coefficient model for ENSO over DSFW, is the best model according to the lowest DIC value. From

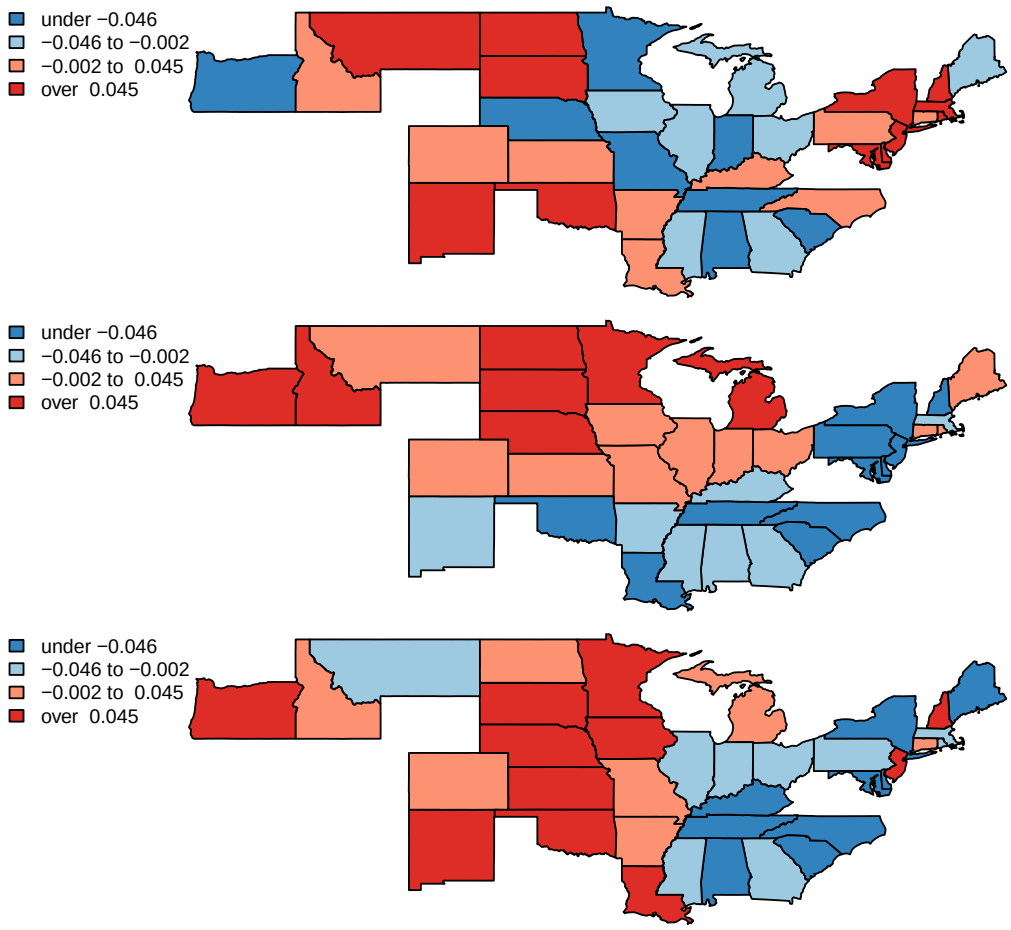

Figure 2.12: *Under SPDE model, the posterior mean of ENSO's coefficients over space at time 34,64,91*

the exploration of figure 2.8 and figure 2.11, we also find that a potential lag effect may exist between ENSO and DSFW. In this study, we add the lag effect to the model 2. The lag 1,2 and 3 coefficients are created by shifting ENSO data by 1 month, 2 months and 3 months. The model 7 can be described by observation equation 2.30 and state equation 2.31 and 2.32. The new fixed effect is the lag effect with length p. We start with length 3 and go though a backward elimination procedure.

Measure	Step 1	Step 2	Step 3
DIC	8766.76	8759.78	8759.38
Main (SD)	1.075* (0.211)	1.150* (0.208)	1.871*(0.100)
Lag 1 (SD)	-1.050* (0.388)	-0.975* (0.392)	1.688*(0.103)
Lag 2 (SD)	0.442 (0.391)	-0.021 (0.226)	
Lag 3 (SD)	-0.297 (0.207)		

Table 2.1: *Backward Elimination of the lag effect in ENSO*

$$\mathbf{y}_t = \Upsilon + z_t\beta_0 + z_{lag(p)}\beta_{lag(p)} + z_t\beta_t + \boldsymbol{\xi}_t + \ \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}), \tag{2.30}$$

$$\boldsymbol{\xi}_t = \alpha_1\boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_t, \ \boldsymbol{\omega}_t \sim PGMRF(\mathbf{0}, \mathbf{W}^{-1}) \tag{2.31}$$

$$\beta_t = \alpha_2\beta_{t-1} + \epsilon_{2t}, \ \epsilon_{2t} \sim N(0, \sigma_{\epsilon_2}^2) \tag{2.32}$$

At the first step, the main effect and all three lags are all put into the model. Table 2.1 demonstrate the posterior mean and standard deviation of the fixed effects. Clearly in step 2, the lag 3 effect is removed and then we fit the model again. In step 3, lag 2 effect is removed as well. Hence, it is determined length of lag is 1 month.

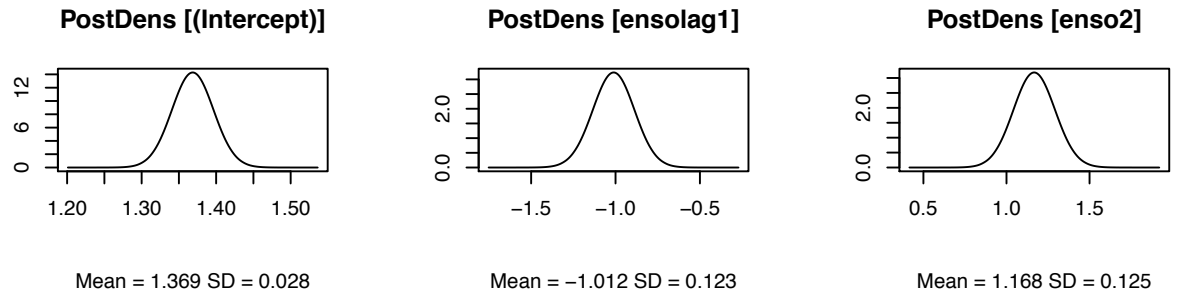

Figure 2.13: *Posterior densities of the fixed effect in the final model*

The posterior densities for the fixed effect of our final model is presented in figure 2.13.

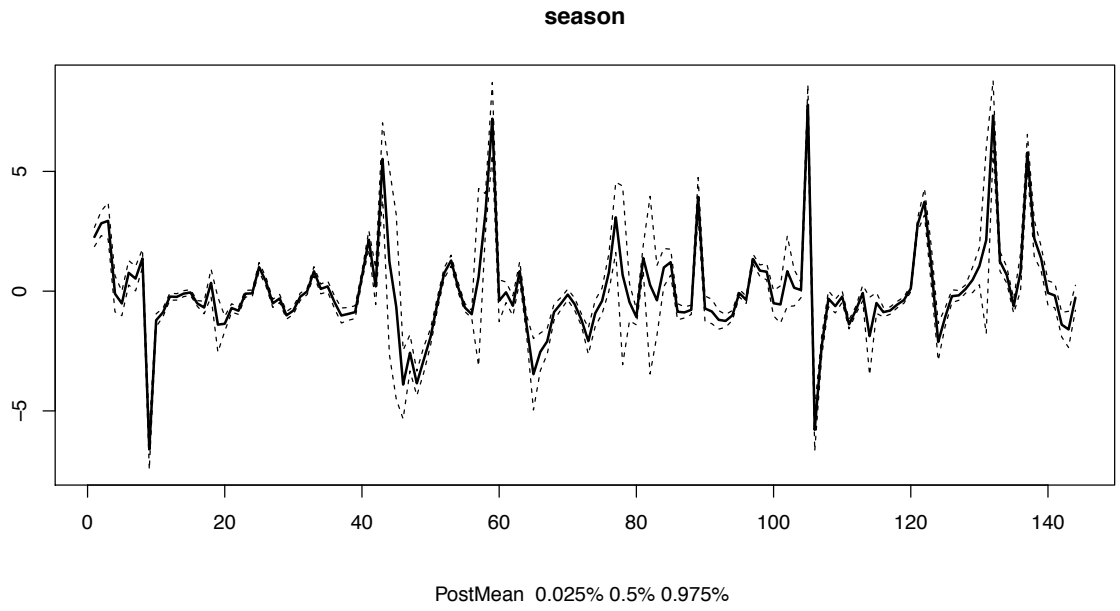

Figure 2.14: *Posterior densities of the time varying coefficients in the final model*

The time-varying coefficients and its posterior credible bond is presented in figure 2.14. All posterior densities of hyper-parameters are given in figure 2.15 .

2.7 Summary, Conclusion and Further Discussion

In this chapter, we demonstrate spatio-temporal modeling of the relationship between DSFW and ENSO using Bayesian approach. The INLA is used for computing posterior inferences. Both discrete spatial domain and continuous spatial domain analysis are demonstrated. One may argue the advantage of one model over another. We use the DIC result which is given for each model as a guideline for our model selection and select model 2 to proceed our inference. We also add a lag effect to evaluate the lagged response from the DSFW. The analysis demonstrate the significance of the lag effect.

The final model's DIC value is 8759.38. The main effects of ENSO shows a significant

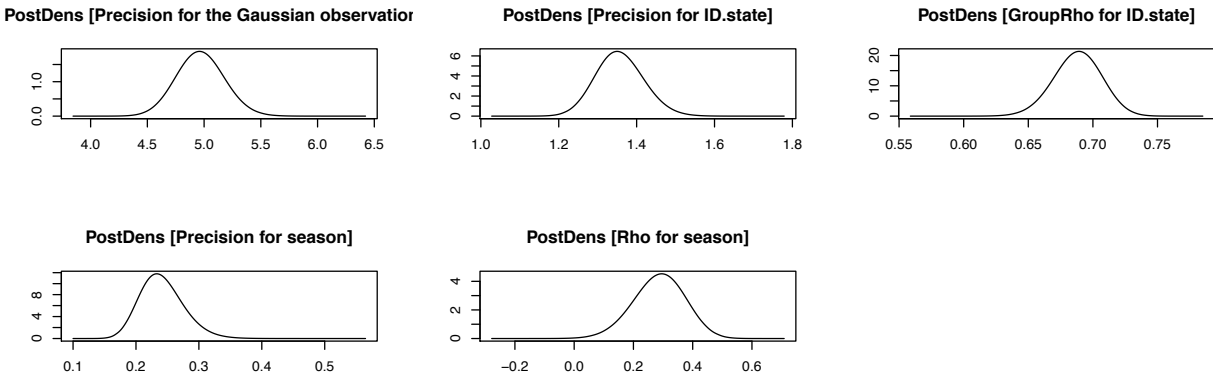

Figure 2.15: *Posterior densities of the hyper parameter in the final model*

positive effect over DSFW, however the true effect of ENSO should combine the main effect with the time-varying effect. This summation determines the positive or negative association and the magnitude at each time points. The respond of DSFW upon the fluctuation of the ENSO could be lagged by 1 month. But, we didn't observe any obvious pattern for the effect of ENSO among the spatial domain of the United States.

Beside making inference for our model, we also did a little comparison between a dynamic models in discrete spatial domain and continuous spatial domain. The continuous spatial domain allows a more flexible neighborhood structure which takes into account for more information than the first order neighborhood structure in the discrete domain. Also, you can put more (less) vertices into the Delaunay triangulation to increase (decrease) resolutions at will. One can also provide a specific prediction at any point, comparing with the prediction using grids in the discrete domain. Even through the advantage of continuous spatial domain model does not show in this particular problem, this framework still weights a lot for future exploration.

Moreover, how to choose if coefficients change over time or over time and space can depend on the DIC value. It can also depend on if the result is interpretable and how much details one wants to put into the model. The advantage of the dynamic models is the

fit may be better than other type of the models. As a result, it can be hard to interpret the results. In our research problem, we can interpret the time-varying coefficients as the effect of ENSO changed over time. In other scenarios, the time-varying coefficients may not be interpretable. In that case, we will not suggest a time-varying coefficients model. Still in practice, one may favor one to another, but we want to provide solutions in multiple ways and allow researchers to tailer those models to the data at hand.

Bibliography

M. A. A. Adviento-Borbe, M. L. Haddix, D. L. Binder, D. T. Walters, and A. Dobermann. Soil greenhouse gas fluxes and global warming potential in four high-yielding maize systems. *Global Change Biology*, 13(9):1972–1988, 2007. ISSN 1365-2486.

Anthony G. Bamston, Muthuvel Chelliah, and Stanley B. Goldenberg. Documentation of a highly enso-related sst region in the equatorial pacific: Research note. *Atmosphere-Ocean*, 35(3):367–383, 1997.

C.F. Baum. Stata tip 63: Modeling proportions . *Stata*, 8(2), 2008.

N. H. Bingham. Positive definite functions on spheres. *Proceedings of the Cambridge Philosophical Society*, 73:145, 1973.

M. Blangiardo and M. Cameletti. *Spatial and Spatio-tempral Bayesian Models with R-INLA*. Wiley, 2016.

S. Bochner. Hilbert distances and positive definite functions. *Annals of Mathematics*, 42 (3):647–656, 1941.

Michela Cameletti, Finn Lindgren, Daniel Simpson, and Håvard Rue. Spatio-temporal modeling of particulate matter concentration through the spde approach. *AStA Advances in Statistical Analysis*, 97(2):109–131, 2013.

N. Cressie. *Statistics for Spatial Data*. Wiley, New York, NY, 1993.

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

Noel Cressie, Sandy Burden, Walter Davis, Pavel N. Krivitsky, Payam Mokhtarian, Thomas Suesse, and Andrew Zammit-Mangion. Capturing multivariate spatial dependence: Model, estimate and then predict. *Statist. Sci.*, 30(2):170–175, 05 2015.

J. Du and C. Ma. Spherically invariant vector random fields in space and time. *IEEE Transactions on Signal Processing*, 59(12):5921–5929, Dec 2011.

Juan Du and Chunsheng Ma. Variogram matrix functions for vector random fields with second-order increments. *Mathematical Geosciences*, 44(4):411–425, 2012.

Juan Du, Chunsheng Ma, and Yang Li. Isotropic variogram matrix functions on spheres. *Mathematical Geosciences*, 45(3):341–357, 2013.

W. Feller. *An introduction to probability theory and its applications*, volume 2. Wiley, New York, NY, 1971.

Gregory Gaspari and Stephen E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554): 723–757, 1999.

Alan E. Gelfand, Hyon-Jung Kim, C. F. Sirmans, and Sudipto Banerjee. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396, 2003.

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

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Tilmann Gneiting, William Kleiber, and Martin Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491):1167–1177, 2010.

T.W. Griffin and J.A. Kelley. Days suitable for fieldwork in arkansas. *University of Arkansas Division of Agriculture FSA33*, 2010. URL http://bit.ly/1o427Cj.

Joseph Guinness and Montserrat Fuentes. Isotropic covariance functions on spheres: Some properties and modeling considerations. *Journal of Multivariate Analysis*, 143:143 – 152, 2016.

Y. Hafez and M. Almazroui. A recent study on the relationship between global radiative forcing and global annual climatic variability. *Atmospheric and Climate Sciences*, 5: 23–55, 2015.

J. Hatfield, G.Takle, R. Grotjahn, P. Holden, R. C. Izaurralde, T. Mader, E. Marshall, and D. Liverman. *Ch. 6: Agriculture. Climate Change Impacts in the United States: The Third National Climate Assessment*. U.S. Global Change Research Program, 2014.

M.J. Heaton, M. Katzfuss, C. Berrett, and D.W. Nychka. Constructing valid spatial processes on the sphere using kernel convolutions. *Environmetrics*, 25(1):2–15, 2014.

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

Hae Kyung IM, Michael L. Stein, and Zhengyuan Zhu. Semiparametric estimation of spectral density with irregular observations. *Journal of the American Statistical Association*, 102(478):726–735, 2007.

Jaehong Jeong and Mikyoung Jun. Covariance models on the surface of a sphere: when does it matter? *Stat*, 4(1):167–182, 2015.

A.G. Journel and Ch. J. Huijbregts. *Mining Geostatistics*. The Blackburn Press, 1978.

Mikyoung Jun and Michael L. Stein. An approach to producing space-time covariance functions on spheres. *Technometrics*, 49(4):468–479, 2007.

Mikyoung Jun and Michael L. Stein. Nonstationary covariance models for global data. *Ann. Appl. Stat.*, 2(4):1271–1289, 12 2008.

Annika Lang and Christoph Schwab. Isotropic gaussian random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations. *Ann. Appl. Probab.*, 25(6):3047–3094, 12 2015.

Yang Li and Zhengyuan Zhu. Modeling nonstationary covariance function with convolution on sphere. *Computational Statistics and Data Analysis*, 104:233 – 246, 2016.

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

C. Ma. Stationary and isotropic vector random fields on spheres. *Math Geosci*, 44:765–778, 2012.

Chunsheng Ma. Isotropic covariance matrix functions on all spheres. *Mathematical Geosciences*, 47(6):699–717, 2015.

Chunsheng Ma. Stochastic representations of isotropic vector random fields on spheres. *Stochastic Analysis and Applications*, 34(3):389–403, 2016.

T.B. Mark, J. D'Antoni, and T.W. Griffin. Revisiting days suitable for fieldwork relative to global climate cycles. In *Southern Agricultural Economics Association Annual Meeting*, Dallas, Texes, 2014a.

T.B. Mark, J.M. D'Antoni, and T.W. Griffin. How do global weather patterns influence days suitable for field work? In *Agricultural and Applied Economics Association Annual Meeting*, 2014b.

T.B. Mark, T.W. Griffin, and J. D'Antoni. Making farm management decisions based on global weather events. In *American Society of Agricultural and Biological Engineers Climate Change Symposium*, Chicago, Illinois, 2015.

Georges Matheron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, 1963. ISSN 0361-0128. doi: 10.2113/gsecongeo.58.8.1246.

H.S. Migon, D. Gamerman, H.F. Lopes, and M.A.R. Ferreira. *Bayesian Dynamic Models in Handbook of statistis, volume 25*. Elsevier:North-Holland, 2005.

S. George Philander. *El Nino, La Nina, and the Southern Oscillation*. Academic Press, 1989.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL http://www.R-project.org/.

Havard Rue and Leonhard Held. *Gaussian Markov Random Fields: Theory And Applications (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC, 2005. ISBN 1584884320.

Havard Rue and Sara Martino. Approximate bayesian inference for hierarchical gaussian markov random field models. *Journal of Statistical Planning and Inference*, 137(10): 3177 – 3192, 2007. ISSN 0378-3758.

Havard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009.

Ramiro Ruiz-Cárdenas, Elias T. Krainski, and Havard Rue. Direct fitting of dynamic models using integrated nested laplace approximations {INLA}. *Computational Statistics and Data Analysis*, 56(6):1808 – 1828, 2012.

S.R. Sain, R. Furrer, and N. Cressie. A spatial analysis of multivariate output from regional climate models. *Ann. Appl. Statist*, 5:150–175, 2011.

RL Schilling, R Song, and Z Vondracek. *Bernstein functions: Theory and applications*. De Gruyter, Berlin, 2012.

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

Michael L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.

Luke Tierney and Joseph B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of The American Statistical Association*, 81:82–86, 03 1986.

Juan C. Vivar and Marco A. R. Ferreira. Spatiotemporal models for gaussian areal data. *Journal of Computational and Graphical Statistics*, 18(3):658–674, 2009.

David Vernon Widder. *The Laplace Transform*. Princeton University Press, Princeton, N. J., 1941.

Appendix A

Appendix

Proof of Theorem 1

Statement (ii) implies (i) because of Theorem 2 of Bochner (1941) and Theorem 3 of Schoenberg (1942). Lemma 2 of Bingham (1973) indicates that (i) implies (ii).

The equivalence between (iii) and (iv) is given by applying $\vartheta = \frac{\pi}{2} - \arcsin(\cos\vartheta)$, so that $\gamma(\vartheta) = \gamma(\frac{\pi}{2} - \arcsin(\cos\vartheta))$. Define $g(x) = \gamma_0 - \gamma\left(\frac{\pi}{2} - \arcsin x\right)$, where $\gamma_0 \geq \max_{0\leq\vartheta\leq\pi} \gamma(\vartheta)$. Then

$$\begin{aligned} g(x) &= \gamma_0 - \gamma\left(\frac{\pi}{2} - \arcsin x\right), \\ g(x) - g(-x) &= \gamma\left(\frac{\pi}{2} - \arcsin x\right) - \gamma\left(\frac{\pi}{2} + \arcsin x\right), \end{aligned}$$

from which the absolute monotonicity of functions in (iii) implies those in (iv), and vice versa.

Next, we will show that (ii) and (iii) are equivalent. First, (iii) can be derived from

(ii) by letting

$$g(x) = \sum_{n=0}^{\infty} b_n x^n, \ x \in [-1, 1],$$

where the right-hand series absolutely converges to a continuous function on [-1, 1], since $\{b_n, n = 0, 1, 2, \ldots\}$ is assumed to be a summable sequence of nonnegative numbers. Clearly, $g(x)$ is absolutely monotone on [0,1]. So is

$$g(x) - g(-x) = \sum_{n=1}^{\infty} (1 - (-1)^n) b_n x^n = 2 \sum_{n=0}^{\infty} b_{2n+1} x^{2n+1}, \ x \in [0, 1].$$

Second, suppose that (iii) holds, $g(x)$ is continuous on [-1,1] and absolutely monotone on [0,1], and that $g(x) - g(-x)$ is absolutely monotone on [0,1]. By Theorem 3a of Widder (1941), page 146, $g(x)$ possesses a Taylor series

$$g(x) = \sum_{n=0}^{\infty} \frac{g^{(n)}(0)}{n!} x^n, \ 0 \le x \le 1, \tag{A.1}$$

with $g^{(n)}(0) \ge 0, n = 0, 1, \ldots$. What remains to prove is that (A.1) holds on the interval $[-1, 0]$ as well, based on which (1.3) is obtained from (1.4) directly. Since $g(x) - g(-x)$ is absolutely monotone on [0,1], it possesses a Taylor series

$$g(x) - g(-x) = \sum_{n=0}^{\infty} \frac{b_n}{n!} x^n, \ 0 \le x \le 1,$$

where

$$b_n = \begin{cases} 0, & n \text{ is zero or even}, \\ \frac{d^n}{dx^n}(g(x) - g(-x))\big|_{x=0} = 2g^{(n)}(0), & n \text{ is odd}. \end{cases}$$

Taking the difference between $g(x)$ and $g(x) - g(-x)$ results in

$$\begin{aligned} g(-x) &= g(x) - \{g(x) - g(-x)\} \\ &= \sum_{n=0}^{\infty} \frac{g^{(n)}(0)}{n!} x^n - 2 \sum_{n=0}^{\infty} \frac{g^{(2n+1)}(0)}{(2n+1)!} x^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{g^{(2n)}(0)}{(2n)!} x^{2n} - \sum_{n=0}^{\infty} \frac{g^{(2n+1)}(0)}{(2n+1)!} x^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{g^{(n)}(0)}{n!} (-x)^n, \ 0 \le x \le 1. \end{aligned}$$

In other words, the Taylor series expansion (A.1) of $g(x)$ holds on the whole interval $[-1, 1]$.

Proof of Theorem 2

(i) It is known that $f_1(f_2(x))$ is absolutely monotone on $[a, b]$, if $f_1(x)$ and $f_2(x)$ are absolutely monotone on $[a, b]$ and $a \le f_2(x) \le b$ (see, e.g., Theorem 2a, page 146 of Widder (1941). Notice that $f_2(x) = \arcsin x$ is absolutely monotone on $[0, 1]$ and $0 \le f_2(x) \le \frac{\pi}{2}, x \in [0, 1]$. As a consequence, the absolute monotonicity of $f_1(x) = \gamma_0 - \gamma\left(\frac{\pi}{2} - x\right)$ implies that of $f_1(f_2(x)) = \gamma_0 - \gamma(\frac{\pi}{2} - \arcsin x)$. Similarly, the absolute monotonicity of $\gamma(\frac{\pi}{2} + \arcsin x) - \gamma(\frac{\pi}{2} - \arcsin x)$ is obtained. By Theorem 1(iv), $\gamma(\vartheta)$ is a variogram on $\mathbb{S}^\infty$.

(ii) Under the completely monotone assumption, $\gamma_0 - \gamma(\vartheta)$ is nonnegative and continuous on $[0, \pi]$, and

$$(-1)^{n+1} \gamma^{(n)}(\vartheta) = (-1)^n \frac{d^n}{d\vartheta^n} \left(\gamma_0 - \gamma(\vartheta)\right), \ n = 1, 2, \ldots,$$

are nonnegative and decreasing on $(0, \pi)$. Thus,

$$(-1)^{n+1}\gamma^{(n)}(x) \geq 0, \ x \in (0, \pi), \tag{A.2}$$

and

$$(-1)^{n+1}\gamma^{(n)}\left(\frac{\pi}{2} - x\right) + (-1)^{n}\gamma^{(n)}\left(\frac{\pi}{2} + x\right) \geq 0, \ x \in \left(0, \frac{\pi}{2}\right), \ n = 1, 2, \ldots. \tag{A.3}$$

Next we apply the above two inequalities to show that the condition in Part (ii) entails that in Part (i).

It follow from (A.2) that

$$\frac{d^n}{dx^n}\left(\gamma_0 - \gamma\left(\frac{\pi}{2} - x\right)\right) = -\frac{d^n}{dx^n}\gamma\left(\frac{\pi}{2} - x\right) = (-1)^{n+1}\gamma^{(n)}\left(\frac{\pi}{2} - x\right) \geq 0, \ x \in (0, 1), \ n = 1, 2, \ldots,$$

i.e., $\gamma_0 - \gamma(\frac{\pi}{2} - x)$ is absolutely monotone on $[0, 1]$. Applying inequality (A.3), we obtain, for an even n,

$$\begin{aligned}
\frac{d^n}{dx^n}\left(\gamma\left(\frac{\pi}{2} + x\right) - \gamma\left(\frac{\pi}{2} - x\right)\right) &= \gamma^{(n)}\left(\frac{\pi}{2} + x\right) - (-1)^{n}\gamma^{(n)}\left(\frac{\pi}{2} - x\right) \\
&= (-1)^{n}\gamma^{(n)}\left(\frac{\pi}{2} + x\right) + (-1)^{n+1}\gamma^{(n)}\left(\frac{\pi}{2} - x\right) \\
&\geq 0,
\end{aligned}$$

and, for an odd n,

$$\begin{aligned}
& \frac{d^n}{dx^n}\left(\gamma\left(\frac{\pi}{2}+x\right)-\gamma\left(\frac{\pi}{2}-x\right)\right) \\
= \; & \gamma^{(n)}\left(\frac{\pi}{2}+x\right)-(-1)^n\gamma^{(n)}\left(\frac{\pi}{2}-x\right) \\
= \; & 2(-1)^{n+1}\gamma^{(n)}\left(\frac{\pi}{2}+x\right)+\left\{(-1)^n\gamma^{(n)}\left(\frac{\pi}{2}+x\right)+(-1)^{n+1}\gamma^{(n)}\left(\frac{\pi}{2}-x\right)\right\} \\
\geq \; & 0.
\end{aligned}$$

The proof is completed by applying Theorem 2 (i).

Proof of Theorem 3

(i) Since (1.10) is a polynomial, both $\gamma\left(\frac{\pi}{2}-x\right)$ and $\gamma\left(\frac{\pi}{2}+x\right)$ are polynomials, and, from the binomial theorem with b_0 set to be 0,

$$\begin{aligned}
\gamma\left(\frac{\pi}{2}-x\right) &= \sum_{k=1}^{p} b_k\left(\frac{\pi}{2}-x\right)^k \\
&= \sum_{k=0}^{p} b_k \sum_{j=0}^{k}(-1)^j\binom{k}{j}\left(\frac{\pi}{2}\right)^{k-j}x^j \\
&= \sum_{k=0}^{p} a_k x^k, \\
\gamma\left(\frac{\pi}{2}+x\right) &= \sum_{k=0}^{p}(-1)^k a_k x^k, \; x\in\left[0,\frac{\pi}{2}\right],
\end{aligned}$$

where

$$a_k=(-1)^k\sum_{j=k}^{p} b_j\binom{j}{k}\left(\frac{\pi}{2}\right)^{j-k}, \; k=0,1,\ldots,p.$$

Take $\gamma_0 \geq a_0 = \gamma(\frac{\pi}{2}) = \sum_{j=1}^{p} b_j \left(\frac{\pi}{2}\right)^j$. Under the assumption (1.11), a_k and $a_k - (-1)^k a_k$ $(k = 1, 2, \ldots, p)$ are nonpositive, so that

$$\gamma_0 - \gamma\left(\frac{\pi}{2} - x\right) \;=\; \gamma_0 - a_0 - \sum_{k=1}^{p} a_k x^k,$$

and

$$\gamma\left(\frac{\pi}{2} + x\right) - \gamma\left(\frac{\pi}{2} - x\right) \;=\; \sum_{k=1}^{p}((-1)^k a_k - a_k)x^k$$

have all coefficients being nonnegative, and thus are absolutely monotone on $[0, \pi]$. By Theorem 2 (i), $\gamma(\vartheta)$ is an isotropic variogram on $\mathbb{S}^\infty$.

(ii) Given an isotropic variogram $\gamma(\vartheta)$ on $\mathbb{S}^\infty$, it follows from Corollary 2 that

$$h_1(x) = \gamma\left(\frac{\pi}{2} + x\right) - \gamma\left(\frac{\pi}{2} - x\right) = \sum_{k=1}^{p}((-1)^k a_k - a_k)x^k$$

is increasing on $\left(0, \frac{\pi}{2}\right]$. In particular, $h_1(x) \geq 0 = h_1(0), x \in \left[0, \frac{\pi}{2}\right]$, from which we obtain

$$0 \leq \lim_{x \to 0_+} \frac{h_1(x) - h_1(0)}{x} = -2a_1.$$

Also, from Corollary 2,

$$h_2(x) = \gamma\left(\frac{\pi}{2} + x\right) + \gamma\left(\frac{\pi}{2} - x\right) = \sum_{k=0}^{p}((-1)^k a_k + a_k)x^k$$

is decreasing on $\left(0, \frac{\pi}{2}\right]$, so that $h_2(x) \leq h_2(0), x \in \left[0, \frac{\pi}{2}\right]$. As a consequence,

$$0 \geq \lim_{x \to 0_+} \frac{h_2(x) - h_2(0)}{x^2} = 2a_2.$$

(iii) This follows directly from Parts (i) and (ii), as a particular case $p = 2$.

Proof of Theorem 4

The equivalence of (i) and (ii) is established by Du et al. (2013). The equivalence of (ii) and (iii) is given by letting $x = \cos\vartheta$ and $\vartheta = \frac{\pi}{2} - \arcsin(\cos\vartheta)$. Thus,

$$\boldsymbol{\gamma}(\vartheta) = \boldsymbol{\gamma}\left(\frac{\pi}{2} - \arcsin x\right) = \sum_{n=1}^{\infty} \mathbf{B}_n(1 - x^n), \ x \in [-1, 1], \tag{A.4}$$

where $\{\mathbf{B}_n, n = 1, 2, \ldots\}$ is a summable sequence of positive definite $m \times m$ matrices.

Proof of Theorem 5

First we need verify that the infinite series of right hand side of (1.21) converges in mean square sense by Cauchy's criterion. For any positive integers n_1, n_2, we have

$$
\begin{aligned}
& E\left(\sum_{i=n_1}^{n_1+n_2} \mathbf{B}_i^{\frac{1}{2}}\mathbf{V}_i\left(\alpha_i^{-1}-\frac{P_i^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_i^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)\right)\left(\sum_{j=n_1}^{n_1+n_2} \mathbf{B}_j^{\frac{1}{2}}\mathbf{V}_j\left(\alpha_j^{-1}-\frac{P_j^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_j^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)\right)' \\
= & E\sum_{i=n_1}^{n_1+n_2}\sum_{j=n_1}^{n_1+n_2} \mathbf{B}_i^{\frac{1}{2}}(\mathbf{V}_i\mathbf{V}_j')(\mathbf{B}_j^{\frac{1}{2}})'\left(\alpha_i^{-1}-\frac{P_i^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_i^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)\left(\alpha_j^{-1}-\frac{P_j^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_j^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right) \\
= & \sum_{i=n_1}^{n_1+n_2}\sum_{i=n_1}^{n_1+n_2} \mathbf{B}_i^{\frac{1}{2}}E(\mathbf{V}_i\mathbf{V}_j')(\mathbf{B}_j^{\frac{1}{2}})'E\left(\alpha_i^{-1}-\frac{P_i^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_i^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)\left(\alpha_j^{-1}-\frac{P_j^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_j^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right) \\
= & \sum_{i=n_1}^{n_1+n_2} \mathbf{B}_i^{\frac{1}{2}}E(\mathbf{V}_i\mathbf{V}_i')(\mathbf{B}_i^{\frac{1}{2}})'E\left(\alpha_i^{-1}-\frac{P_i^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_i^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)\left(\alpha_i^{-1}-\frac{P_i^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_i^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right) \\
= & \sum_{i=n_1}^{n_1+n_2} \alpha_i^2\mathbf{B}_i E\left(\alpha_i^{-2}+\frac{(P_i^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U}))(P_i^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U}))'}{P_i^{\frac{d-1}{2}}(1)}-2\alpha_i^{-1}\frac{P_i^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_i^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right) \\
= & \sum_{i=n_1}^{n_1+n_2} \mathbf{B}_i+\sum_{i=n_1}^{n_1+n_2}\alpha_i^2\mathbf{B}_i\frac{(P_i^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{x}))}{\alpha^2 P_i^{\frac{d-1}{2}}(1)}-0 \\
= & 2\sum_{i=n_1}^{n_1+n_2}\mathbf{B}_i \\
\to & 0,\ n_1, n_2\to\infty,
\end{aligned}
$$

where the second equality holds because $\{\mathbf{V}_n, n\in\mathbb{N}\}$ and $\mathbf{U}$ are independent, the third and fifth equalities follow from Lemmas 2 and 3 of Ma (2016), and the last one is due to the convergent assumption of $\sum_{n=1}^{\infty}\mathbf{B}_n$. Hence $\mathbf{Z}(\mathbf{x})$ in (1.21) is well defined. Now the variogram matrix function of $\{\mathbf{Z}(\mathbf{x}), \mathbf{x}\in\mathbb{S}^d\}$ is given by

$$
\begin{aligned}
\boldsymbol{\gamma}(\mathbf{x}_1,\mathbf{x}_2) &= \frac{1}{2}E(\mathbf{Z}(\mathbf{x}_1)-\mathbf{Z}(\mathbf{x}_2)-E(\mathbf{Z}(\mathbf{x}_1)-\mathbf{Z}(\mathbf{x}_2)))(\mathbf{Z}(\mathbf{x}_1)-\mathbf{Z}(\mathbf{x}_2)-E(\mathbf{Z}(\mathbf{x}_1)-\mathbf{Z}(\mathbf{x}_2)))' \\
&= \frac{1}{2}E\left(\sum_{n=1}^{\infty}\mathbf{B}_n^{\frac{1}{2}}\mathbf{V}_n\left(\alpha_n^{-1}-\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_n^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)-\sum_{n=0}^{\infty}\mathbf{B}_n^{\frac{1}{2}}\mathbf{V}_n\left(\alpha_n^{-1}-\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_n^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)\right) \\
&\quad \left(\sum_{n=0}^{\infty}\mathbf{B}_n^{\frac{1}{2}}\mathbf{V}_n\left(\alpha_n^{-1}-\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_n^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)-\sum_{n=0}^{\infty}\mathbf{B}_n^{\frac{1}{2}}\mathbf{V}_n\left(\alpha_n^{-1}-\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}'\mathbf{U})}{(P_n^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)\right)' \\
&= \frac{1}{2}E\left(\sum_{n=0}^{\infty}\mathbf{B}_n^{\frac{1}{2}}\mathbf{V}_n\left(\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}_2'\mathbf{U})}{(P_n^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}-\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}_1'\mathbf{U})}{(P_n^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)\right)\left(\sum_{n=0}^{\infty}\mathbf{B}_n^{\frac{1}{2}}\mathbf{V}_n\left(\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}_2'\mathbf{U})}{(P_n^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}-\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}_1'\mathbf{U})}{(P_n^{\frac{d-1}{2}}(1))^{\frac{1}{2}}}\right)\right)' \\
&= \frac{1}{2}\sum_{n=0}^{\infty}\mathbf{B}_n^{\frac{1}{2}}E(\mathbf{V}_n\mathbf{V}_n')(\mathbf{B}_n^{\frac{1}{2}})'(P_n^{\frac{d-1}{2}}(1))^{-1}E[(P_n^{\frac{d-1}{2}}(\mathbf{x}_2'\mathbf{U})-P_n^{\frac{d-1}{2}}(\mathbf{x}_1'\mathbf{U}))(P_n^{\frac{d-1}{2}}(\mathbf{x}_2'\mathbf{U})-P_n^{\frac{d-1}{2}}(\mathbf{x}_1'\mathbf{U}))] \\
&= \frac{1}{2}\sum_{n=0}^{\infty}\alpha_n^2\mathbf{B}_n(P_n^{\frac{d-1}{2}}(1))^{-1}\left(\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}_2'\mathbf{x}_2)}{\alpha_n^2}+\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}_1'\mathbf{x}_1)}{\alpha_n^2}-2\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}_2'\mathbf{x}_1)}{\alpha_n^2}\right) \\
&= \sum_{n=0}^{\infty}\mathbf{B}_n\left(1-\frac{P_n^{\frac{d-1}{2}}(\mathbf{x}_2'\mathbf{x}_1)}{P_n^{\frac{d-1}{2}}(1)}\right) \\
&= \sum_{n=0}^{\infty}\mathbf{B}_n(1-p_n^{(\frac{d-1}{2})}(\cos\vartheta(\mathbf{x}_1,\mathbf{x}_2))),
\end{aligned}
$$

where the first two equalities follow because $\mathbf{Z}_0$ cancels out and $E(\mathbf{Z}(\mathbf{x}_1)-\mathbf{Z}(\mathbf{x}_2))$ is zero. The next two equalities hold due to independence of $\{\mathbf{V}_n, n\in\mathbb{N}_0\}$ and $\mathbf{U}$. Then the last three equalities result from the properties of ultraspherical polynomials specified in lemma 2 of Ma (2016). Actually the variogram matrix (1.22) is the ultraspherical polynomial expansion of a valid variogram on all spheres by equation (6) and (10) in Du et al. (2013).

Proof of Proposition 1

Let's denote $\pi(\cdot|\cdot)$ as the conditional density of its arguments. The common objective of Bayesian inference is the posterior distribution. This posterior distribution can be shown as

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) \prod_{i \in I} \pi(y_i|x_i, \boldsymbol{\theta}) \tag{A.5}$$

$$\propto \pi(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{n/2} exp(-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_{i \in I} log\pi(y_i|x_i, \boldsymbol{\theta})) \tag{A.6}$$

Prior to introduce the Integrated Nested Laplace Approximation, we introduce the Gaussian Approximation first. In this setting, $exp(-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_{i \in I} log\pi(y_i|x_i, \boldsymbol{\theta}))$ is the primary part to be approximated. Matching the mode and curvature at the mode, the resulting approximation is

$$\tilde{\pi}(x|\boldsymbol{\theta}, \mathbf{y}) \propto exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T(\mathbf{Q} + diag(\mathbf{c}))(\mathbf{x} - \boldsymbol{\mu})) \tag{A.7}$$

where $\boldsymbol{\mu}$ is the mode of $\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. Note that both $\boldsymbol{\mu}$ and $\mathbf{Q}$ depend on $\boldsymbol{\theta}$. The terms of vector $\mathbf{c}$ are due to the second order terms in the Taylor expansion of $\sum log\pi(y_i|x_i)$ at the modal value $\boldsymbol{\mu}$ and these terms are zero for approximation.

In those cases where dimension of $\boldsymbol{\theta}$ is small, it is possible to derive an independence sampler by reusing equation A.7 to build an approximation of the marginal posterior for $\boldsymbol{\theta}$.

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \int C\pi(\boldsymbol{\theta}, \mathbf{x}|\mathbf{y})d\mathbf{x} \tag{A.8}$$

$$= C\int \pi(\boldsymbol{\theta}, \mathbf{x}, \mathbf{y})d\mathbf{x} \tag{A.9}$$

$$= C\int \pi(\boldsymbol{\theta}, \mathbf{x})\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})d\mathbf{x} \tag{A.10}$$

$$= C\int \pi(\boldsymbol{\theta}, \mathbf{x})exp(log\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}))d\mathbf{x} \tag{A.11}$$

Laplace approximation can be applied to function of $\mathbf{x}$ in the integrand above with fixed $\boldsymbol{\theta}$ as in equation (4.1) in Tierney and B. Kadane (1986). Then marginal poster density of $\boldsymbol{\theta}$ can be written as

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = Cdet|\mathbf{Q}(\boldsymbol{\theta})|^{1/2}\pi(\boldsymbol{\theta}, \mathbf{x}*)exp(log\pi(\mathbf{y}|\boldsymbol{\theta}, \mathbf{x}*)) \tag{A.12}$$

where $x*$ is the mode of the full conditional for x for a given $\boldsymbol{\theta}$. This is equivalent to equation (3) at Rue et al. (2009) since $\tilde{\pi}_G(x|\boldsymbol{\theta}, \mathbf{y})$ is the normalizing constant of equation A.12. Based on this idea, the INLA is developed as the section 2.2.3 demonstrated.

Posterior Distribution of Model 3

At first, the hyper-parameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'$ is split into $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ which are the parameters involved in each individual state equation. Assume two state equations are independent with each other, the posterior distribution can be derived as

$$\pi(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta}|\mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\xi}, \boldsymbol{\beta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\pi(\boldsymbol{\xi}|\boldsymbol{\theta}_1)\pi(\boldsymbol{\beta}|\boldsymbol{\theta}_2)\pi(\boldsymbol{\theta}1)\pi(\boldsymbol{\theta}_2) \tag{A.13}$$

where $\pi(\cdot)$ denote the probability density function, $y = \{\mathbf{y}_t\}, \boldsymbol{\xi} = \{\boldsymbol{\xi}_t\}, \boldsymbol{\beta} = \{\boldsymbol{\beta}_t\}$ with $t = 1, \ldots, T = 144$. The number of dimension at each time point is $d = 36$. The prior distribution is assumed to be independent as well. The equation A.13 can be written as

$$\begin{aligned}\pi(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta}|\mathbf{y}) \quad \propto \quad & (\prod_{t=1}^{T} \pi(\mathbf{y}_t|\boldsymbol{\xi}_t, \boldsymbol{\beta}_t, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2))(\pi(\boldsymbol{\xi}_1|\boldsymbol{\theta}_1) \prod_{t=2}^{T} \pi(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}, \boldsymbol{\theta}_1)) \\ & \times(\pi(\boldsymbol{\beta}_1|\boldsymbol{\theta}_2) \prod_{t=2}^{T} \pi(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \boldsymbol{\theta}_2))\pi(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2)\end{aligned}$$

From equation 2.27 ,2.19 and 2.20, it follows immediately that the join posterior distribution is given by

$$\begin{aligned}\pi(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta}|\mathbf{y}) \quad \propto \quad & (\sigma_\epsilon^2)^{-\frac{dT}{2}} exp(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^{T} (\mathbf{y}_t - \Upsilon - z_t\boldsymbol{\beta}_t - \boldsymbol{\xi}_t)'(\mathbf{y}_t - \Upsilon - z_t\boldsymbol{\beta}_t - \boldsymbol{\xi}_t)) \\ & \times(\frac{\sigma_{\omega_1}^2}{1-\alpha_1^2})^{-d/2}|W|^{1/2} exp(-\frac{1-\alpha_1^2}{2\sigma_{\omega_1}^2}\boldsymbol{\xi}_1' W \boldsymbol{\xi}_1) \\ & \times(\sigma_{\omega_1}^2)^{d(T-1)/2}|W|^{(T-1)/2} \\ & \times exp(-\frac{1}{2\sigma_{\omega_1}^2} \sum_{t=2}^{T} (\boldsymbol{\xi}_t - \alpha_1\boldsymbol{\xi}_{t-1})' W (\boldsymbol{\xi}_t - \alpha_1\boldsymbol{\xi}_{t-1})) \\ & \times(\frac{\sigma_{\omega_2}^2}{1-\alpha_2^2})^{-d/2}|W|^{1/2} exp(-\frac{1-\alpha_2^2}{2\sigma_{\omega_2}^2}\boldsymbol{\beta}_1' W \boldsymbol{\beta}_1) \\ & \times(\sigma_{\omega_2}^2)^{d(T-1)/2}|W|^{(T-1)/2} \\ & \times exp(-\frac{1}{2\sigma_{\omega_2}^2} \sum_{t=2}^{T} (\boldsymbol{\beta}_t - \beta_0 - \alpha_2\boldsymbol{\beta}_{t-1})' W (\boldsymbol{\beta}_t - \beta_0 - \alpha_2\boldsymbol{\beta}_{t-1})) \\ & \times \pi(\sigma_{\omega_1}^2)\pi(\sigma_{\omega_2}^2)\pi(\alpha_1)\pi(\alpha_2)\end{aligned}$$