MULTIVARIATE GAUSSIAN RANDOM FIELDS: EXTREME VALUES, PARAMETER ESTIMATION AND PREDICTION

By

Yuzhen Zhou

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics – Doctor of Philosophy

2015

ABSTRACT

MULTIVARIATE GAUSSIAN RANDOM FIELDS: EXTREME VALUES, PARAMETER ESTIMATION AND PREDICTION

By

Yuzhen Zhou

Motivated by the wide applications of multivariate Gaussian random fields in spatial modeling, we study the tail probability of the extremes, the inference of fractal indices and large covariance modeling of multivariate Gaussian random fields. First, we establish the precise asymptotics for the extremes of bivariate Gaussian random fields by applying the double sum method. The main results can be applied to bivariate Matérn fields. Second, we study the joint asymptotic properties of estimating the fractal indices of bivariate Gaussian random processes under infill asymptotics, which indicates that the estimators are asymptotically independent of the cross correlation in most cases. Third, we define a framework to couple high-dimensional and spatially indexed LiDAR signals with forest variables using a fully Bayesian functional spatial data analysis, which is able to capture within and among LiDAR signal/forest variables association within and across locations. The proposed modeling framework is illustrated by a simulated study and by analyzing LiDAR and spatially coinciding forest inventory data collected on the Penobscot Experimental Forest, Maine.

Copyright by
YUZHEN ZHOU
2015

I dedicate this dissertation to my family.

ACKNOWLEDGMENTS

This is a pleasure to express my sincere gratitude to my advisor Professor Yimin Xiao for his great mentoring and continuous support. I am fully indebted to him for his understanding, patience and encouragement. He is always helpful and tries his best to create opportunities for young researchers. He is open-minded and has great enthusiasm on pursuing excellence in research, which inspires me to set up a career in academia.

I would like to thank Professor Andrew Finley, who is the mentor of the last chapter of the thesis. Andy brings me into a very interesting applied topic. He always shows his enthusiasm in research and gives me many valuable suggestions during the project. I really appreciate all his efforts for helping me make progress.

I also wish to thank Professor Pingshou Zhong. We've been working on a project during the early years of my PhD. His knowledge, experience and encouragement helped me a lot while I was starting to do research.

In addition, I would like to thank Professor Chae Young Lim and Professor Lifeng Luo for serving on my dissertation committee. I am extremely grateful for their assistance and suggestions throughout the thesis work.

Finally, I would like to thank my family and my friends for their love and support. I give my special thanks to my love Yingjie, who was always standing by me in my hard times during this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Multivariate Gaussian random fields

Before presenting the motivation and main work of this thesis, I would like to give the formal definitions of random fields, real valued Gaussian random fields and multivariate Gaussian random fields (see, e.g., [AT07]).

A random field is a stochastic process indexed by a parameter space, which could be a subset of Euclidean space $\mathbb{R}^N$. Formally, it is defined by

Definition 1.1.1 (Random fields)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and T be a topological space. Denote by $\mathbb{R}^T$ be the space of all real-valued function on T. Then, a measurable mapping $X : \Omega \to \mathbb{R}^T$ is called a **real-valued random field**. Measurable mappings from Ω to $(\mathbb{R}^T)^d, d > 1$, are called **multivariate random fields** or vector-valued random fields.

Hence, $X(\omega)$ is a univariate (or multivariate) function and $X(\omega, t)$ its value at t. We usually omit ω and write the random field at $t \in T$ as $X(t)$.

A real-valued Gaussian random field is a random field X indexed by a parameter space T whose finite dimensional distributions of $(X(t_1), ..., X(t_n))^\top$ are multivariate Gaussian for each $n \in \mathbb{N}$ and each $(t_1, ..., t_n) \in T^n$. The distribution of X is determined by its mean and covariance functions, that are

$$\mu(t) := \mathbb{E}(X(t)),\ C(s,t) := \mathrm{Cov(X(s), X(t))}.$$

Let $\mathbf{X}$ be the vector-valued random fields taking values in $\mathbb{R}^d$. Denote by $\mathbf{X} = (X_1, ..., X_d)^\top$ where X_i is its ith coordinate process. $\mathbf{X}$ is called multivariate Gaussian random fields if for any vector $\alpha \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, $\sum_{i=1}^d \alpha_i X_i(t)$ is a real-valued Gaussian random field. The distribution of $\mathbf{X}$ is determined by its vector-valued mean and matrix-valued covariance functions, that are $\mu(t) := (\mathbb{E}(X_1(t)), \mathbb{E}(X_2(t)), ..., \mathbb{E}(X_d(t)))^\top$ and

$$\mathrm{Cov}(\mathbf{X}(s), \mathbf{X}(t)) = \begin{pmatrix} C_{11}(s,t) & C_{12}(s,t) & \cdots & C_{1d}(s,t) \\ C_{21}(s,t) & C_{22}(s,t) & \cdots & C_{2d}(s,t) \\ \vdots & \vdots & \ddots & \vdots \\ C_{d1}(s,t) & C_{d2}(s,t) & \cdots & C_{dd}(s,t) \end{pmatrix},$$

where $C_{ij}(s,t) := \mathrm{Cov}(X_i(s), X_j(t)),\ i, j = 1, 2, ..., d.$

1.2 Overview

This work is motivated by the factor that there is an increasing need for analyzing multivariate spatial datasets [GDFG10, Wac03]. There is a rich literature on modeling univariate spatial data [Cre93, Ste99]. However, in the multivariate setting, model specification is more challenging because we also wish to capture cross-covariance among outcomes and sites [GDFG10, CW11, BCG14].

Multivariate Gaussian random fields are a good candidate model to characterize the covariance structure of the multivariate spatial datasets. [GKS10] introduced the full bivariate Matérn field $\mathbf{X}(t) = (X_1(t), X_2(t))$, which is a $\mathbb{R}^2$-valued, stationary Gaussian random field on $\mathbb{R}^N$ with zero mean and matrix-valued Matérn covariance functions. As spatially correlated error field, this model was applied to probabilistic weather field forecasting for surface pressure and temperature over the North American Pacific Northwest.

While the multivariate Gaussian random fields are widely used in spatial modeling, it raises many interesting problems in the aspects of both theory and modeling. In this work, we focus on three topics in this area: the tail probability of the extremes, the joint asymptotics of fractal indices under infill asymptotics, and large covariance modeling with Gaussian predictive processes. The rest chapters show the details of these three problems.

In Chapter 2, we study the tail probability of the extremes for a class of bivariate spatial model, i.e., $\mathbb{P}\big(\max_{s\in A_1} X_1(s) > u, \max_{t\in A_2} X_2(t) > u\big), \quad \text{as } u \to \infty$. Applying the double sum method [Pit96, Ans06], we establish an explicit form for the tail probability of double extremes for the bivariate field. We found that the area where the cross correlation attains its maximum has highest chance to cause extreme events. Also, the smoothness of the surface for each component affects extreme probability.

In Chapter 3, we study the joint asymptotics of the fractal indices for bivariate Gaussian random processes. We want to see how the cross dependence structure would affect the efficiency of the estimators. The fractal index of each component is estimated respectively by the increment-based method [CW00, CW04]. We established the joint asymptotics of the bivariate estimators under infill asymptotics, which indicated that the estimators are asymptotically independent of the cross correlation in most cases.

In Chapter 4, we define a framework to couple high-dimensional and spatially indexed LiDAR signals with forest variables using a fully Bayesian functional spatial data analysis.

This modeling framework allows us to capture within and among LiDAR signal/forest variables association within and across locations. However, the computational complexity of such models increases in cubic order with the number of spatial locations and the dimension of the LiDAR signal, and the number of forest variables—a characteristic common to multivariate spatial process models. To address this computation challenge, we proposed an approximated model by employing the modified Gaussian predictive processes [BGFS08, FSBG09] twice, both in locations and in heights.

We end the introduction with some notation. For any $t \in \mathbb{R}^N$, $|t|$ denotes its l^2-norm. An integer vector $\mathbf{k} \in \mathbb{Z}^N$ is written as $\mathbf{k} = (k_1, ..., k_N)$. For $\mathbf{k} \in \mathbb{Z}^N$ and $T \in \mathbb{R}_+ = [0, \infty)$, we define the cube $[\mathbf{k}T, (\mathbf{k}+1)T] := \prod_{i=1}^{N}[k_iT, (k_i+1)T]$. For any integer n, $mes_n(\cdot)$ denotes the n-dimensional Lebesgue measure. An unspecified positive and finite constant will be denoted by C_0. More specific constants are numbered by $C_1, C_2, \ldots$.

Chapter 2

Tail asymptotics for the Extremes of Bivariate Gaussian Random Fields

Let $\{\mathbf{X}(t) = (X_1(t), X_2(t))^\top,\ t \in \mathbb{R}^N\}$ be an $\mathbb{R}^2$-valued continuous locally stationary Gaussian random field with $\mathbb{E}[\mathbf{X}(t)] = \mathbf{0}$. For any compact sets $A_1, A_2 \subset \mathbb{R}^N$, precise asymptotic behavior of the excursion probability

$$\mathbb{P}\left(\max_{s\in A_1} X_1(s) > u,\ \max_{t\in A_2} X_2(t) > u\right), \quad \text{as} \quad u \to \infty$$

is investigated by applying the double sum method. The explicit results depend not only on the smoothness parameters of the coordinate fields X_1 and X_2, but also on their maximum correlation ρ.

2.1 Introduction

For a real-valued Gaussian random field $X = \{\mathbf{X}(t),\ t \in T\}$, where T is the parameter set, defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the excursion probability $\mathbb{P}\{\sup_{t\in T} X(t) > u\}$ has been studied extensively. Extending the seminal work of [Pic69], [Pit96] developed a systematic theory on asymptotics of the aforementioned excursion probability for a broad class of Gaussian random fields. Their method, which is called the double sum method, has been further extended by [CL06] to non-Gaussian random fields and, recently, by [DHJ14] to a non-stationary Gaussian random field $\{X(s,t),(s,t) \in \mathbb{R}^2\}$ whose variance function attains its maximum on a finite number of disjoint line segments. For smooth Gaussian random fields, more accurate approximation results have been established by using integral and differential-geometric methods (see, e.g., [Adl00], [AT07], [AW09] and the references therein). For Gaussian and asymptotically Gaussian random fields, the change of measure method was developed by [NSY08] and [Yak13]. Many of the results in the aforementioned references have found important applications in statistics and other scientific areas. We refer to [ATW10] and [Yak13] for further information.

However, only a few authors have studied the excursion probability of multivariate random fields. [PS05] and [DKMR10] established large deviation results for the excursion probability in multivariate case. [Ans06] obtained precise asymptotics for a special class of non-stationary bivariate Gaussian processes, under quite restrictive conditions. [HJ14] recently derived precise asymptotics for the excursion probability of a bivariate fractional Brownian motion with constant cross correlation. The last two papers only consider multivariate processes on the real line $\mathbb{R}$ with specific cross dependence structures. [CX14] established a precise approximation to the excursion probability by using the mean Euler characteristics of the excursion set for a broad class of smooth bivariate Gaussian random fields on $\mathbb{R}^N$. In the present chapter we investigate asymptotics of the excursion probability of non-smooth bivariate Gaussian random fields on $\mathbb{R}^N$, where the methods are totally different from the smooth case.

Our work is also motivated by the recent increasing interest in using multivariate random fields for modeling multivariate measurements obtained at spatial locations (see, e.g., [GDFG10], [Wac03]). Several classes of multivariate spatial models have been introduced by [GKS10], [AGS12] and [KN12]. We will show in Section 2 that the main results of this chapter are applicable to bivariate Gaussian random fields with Matérn cross-covariances introduced by [GKS10]. Furthermore, we expect that the excursion probabilities considered in this chapter will have interesting statistical applications.

Let $\{\mathbf{X}(t), t \in \mathbb{R}^N\}$ be an $\mathbb{R}^2$-valued (not-necessarily stationary) Gaussian random field with $\mathbb{E}[\mathbf{X}(t)] = \mathbf{0}$. We write $\mathbf{X}(t) \triangleq (X_1(t), X_2(t))^\top$ and define

$$r_{ij}(s,t) := \mathbb{E}[X_i(s)X_j(t)], \ i,j = 1,2. \tag{2.1.1}$$

Throughout this chapter, we impose the following assumptions.

i) $r_{ii}(s,t) = 1 - c_i|t-s|^{\alpha_i} + o(|t-s|^{\alpha_i})$, where $\alpha_i \in (0,2)$ and $c_i > 0$ $(i = 1,2)$ are constants.

ii) $|r_{ii}(s,t)| < 1$ for all $|t-s| > 0$, $i = 1,2$.

iii) $r_{12}(s,t) = r_{21}(s,t) := r(|t-s|)$. Namely, the cross correlation is isotropic.

iv) The function $r(\cdot) : [0,\infty) \to \mathbb{R}$ attains maximum only at zero with $r(0) = \rho \in (0,1)$, i.e., $|r(t)| < \rho$ for all $t > 0$. Moreover, we assume $r'(0) = 0, r''(0) < 0$ and there exists $\eta > 0$, for any $s \in [0,\eta]$, $r''(s)$ exists and continuous.

The cross correlation defined here is meaningful and common in spatial statistics where it is usually assumed that the correlation decreases as the distance between two observations increases (see, e.g., [GDFG10], [GKS10]). We only assume that the cross correlation is twice continuously differentiable around the area where the maximum correlation is attained, which is a weaker assumption than that in [CX14] who considered smooth bivariate Gaussian fields.

For any compact sets $A_1, A_2 \subset \mathbb{R}^N$, we investigate the asymptotic behavior of the following excursion probability

$$\mathbb{P}\biggl(\max_{s\in A_1} X_1(s) > u,\ \max_{t\in A_2} X_2(t) > u\biggr), \quad \text{as} \ \ u \to \infty. \tag{2.1.2}$$

The main results of this chapter are Theorems 2.1 and 2.2 below, which demonstrate that the excursion probability (2.1.2) depends not only on the smoothness parameters of the coordinate fields X_1 and X_2, but also on their maximum correlation ρ. The proofs of our Theorems 2.1 and 2.2 will be based on the double sum method. Compared with the earlier works of [LP00], [Ans06] and [HJ14], the main difficulty in the present work is that the correlation function of X_1 and X_2 attains its maximum over the set $D := \{(s, s) : s \in A_1 \cap A_2\}$ which may have different geometric configurations. Several non-trivial modifications for carrying out the arguments in the double sum method have to be made.

This work raises several open questions. For example it would be interesting to study the excursion probabilities when $\{\mathbf{X}(t), t \in \mathbb{R}^N\}$ is anisotropic or non-stationary, or taking values in $\mathbb{R}^d$ with $d \geq 3$. In the last problem, the covariance and cross-covariance structures become more complicated. We expect that the pairwise maximum cross correlations and the size (e.g., the Lebesgue measure) of the set where all the pairwise cross correlations attain their maximum values (if not empty) will play an important role.

The rest sections in this chapter are organized as follows. Section 2.2 states the main theorems with some discussions. We provides an application of the main theorems to the bivariate Gaussian fields with Matérn cross-covariances introduced by [GKS10] in Section 2.3. We state the key lemmas and provide proofs of our main theorems in Section 2.4. The proofs of the lemmas are given in Section 2.5.

2.2 Main Results and Discussions

We recall the Pickands constant first (see, e.g., [Pic69, Pit96]). Let $\chi = \{\chi(t), t \in \mathbb{R}^N\}$ be a (rescaled) fractional Brownian motion with Hurst index $\alpha/2 \in (0,1)$, which is a centered Gaussian field with covariance function $\mathbb{E}[\chi(t)\chi(s)] = |t|^\alpha + |s|^\alpha - |t-s|^\alpha$.

As in [LP00] and [Ans06], we define for any compact sets $\mathbb{S}, \mathbb{T} \subset \mathbb{R}^N$,

$$H_\alpha(\mathbb{S}, \mathbb{T}) := \int_0^\infty e^s \cdot \mathbb{P}\Big(\sup_{t\in\mathbb{S}} \big(\chi(t) - |t|^\alpha\big) > s, \sup_{t\in\mathbb{T}} \big(\chi(t) - |t|^\alpha\big) > s \Big)\, ds. \tag{2.2.1}$$

Let $H_\alpha(\mathbb{T}) = H_\alpha(\mathbb{T}, \mathbb{T})$. Then, the Pickands constant is defined as

$$H_\alpha := \lim_{T\to\infty} \frac{H_\alpha([0,T]^N)}{T^N}, \tag{2.2.2}$$

which is positive and finite (cf. [Pit96]).

Before moving to the tail probability of extremes of a bivariate Gaussian random field, let us consider the tail probability of a standard bivariate Gaussian vector (ξ, η) with correlation ρ. It is known that (see, e.g., [LP00])

$$\mathbb{P}(\xi > u, \eta > u) = \Psi(u,\rho)(1 + o(1)), \text{ as } u \to \infty,$$

where

$$\Psi(u,\rho) := \frac{(1+\rho)^2}{2\pi u^2 \sqrt{1-\rho^2}} \exp\left(-\frac{u^2}{1+\rho} \right).$$

The exponential part of the tail probability above is determined by the correlation ρ. As shown by Theorems 2.2.1 and 2.2.2 below, similar phenomenon also happens for the tail probability of double extremes of $\{\mathbf{X}(t), t \in \mathbb{R}^N\}$, where the exponential part is determined by the maximum cross correlation of the coordinate fields X_1 and X_2.

We will study double extremes of X on the domain $A_1 \times A_2$ where A_1, A_2 are bounded Jordan measurable sets in $\mathbb{R}^N$. That is, the boundaries of A_1 and A_2 have N-dimensional Lebesgue measure 0 (see, e.g., [Pit96], p.105). We only consider the case when $A_1 \cap A_2 \neq \emptyset$, in which the maximum cross correlation ρ can be attained.

If $mes_N(A_1 \cap A_2) \neq 0$, we have the following theorem.

Theorem 2.2.1. *Let $\{\mathbf{X}(t), t \in \mathbb{R}^N\}$ be a bivariate Gaussian random field that satisfies the assumptions in Section 2.1. If $mes_N(A_1 \cap A_2) \neq 0$, then as $u \to \infty$,*

$$\begin{aligned}
&\mathbb{P}\Big(\max_{s\in A_1} X_1(s) > u, \max_{t\in A_2} X_2(t) > u\Big) \\
&= (2\pi)^{\frac{N}{2}}(-r''(0))^{-\frac{N}{2}} c_1^{\frac{N}{\alpha_1}} c_2^{\frac{N}{\alpha_2}} mes_N(A_1 \cap A_2) H_{\alpha_1} H_{\alpha_2} \\
&\quad \times (1+\rho)^{-N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)} u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)} \Psi(u,\rho)(1+o(1)).
\end{aligned} \tag{2.2.3}$$

If $mes_N(A_1 \cap A_2) = 0$, the above theorem is not informative. We have not been able to obtain a general explicit formula in the general. Instead, we consider the special cases

$$A_1 = A_{1,M} \times \prod_{j=M+1}^{N} [S_j, T_j] \quad \text{and} \quad A_2 = A_{2,M} \times \prod_{M+1}^{N} [T_j, R_j], \tag{2.2.4}$$

where $A_{1,M}$ and $A_{2,M}$ are M dimensional Jordan sets with $mes_M(A_{1,M} \cap A_{2,M}) \neq 0$ and $S_j \le T_j \le R_j$, $j = M+1, \dots, N$, $0 \le M \le N-1$. For simplicity of notation, let $mes_0(\cdot) \equiv 1$. Our next theorem shows that the excursion probability is smaller than that in (2.2.3) by a factor of u^{M-N}.

Theorem 2.2.2. *Let $\{\mathbf{X}(t), t \in \mathbb{R}^N\}$ be a bivariate Gaussian random field that satisfies the* *assumptions in Section 2.1, and let A_1, A_2 be as in (2.2.4) with $mes_M(A_{1,M} \cap A_{2,M}) > 0$. Then as $u \to \infty$,*

$$\begin{aligned}
&\mathbb{P}\left(\max_{s\in A_1} X_1(s) > u,\ \max_{t\in A_2} X_2(t) > u\right) \\
&= (2\pi)^{\frac{M}{2}}(-r''(0))^{-\frac{2N-M}{2}} c_1^{\frac{N}{\alpha_1}} c_2^{\frac{N}{\alpha_2}} H_{\alpha_1} H_{\alpha_2} mes_M(A_{1,M} \cap A_{2,M}) \\
&\quad \times (1+\rho)^{2N-M-\frac{2N}{\alpha_1}-\frac{2N}{\alpha_2}} u^{M+N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-2)} \Psi(u,\rho)(1+o(1)).
\end{aligned} \tag{2.2.5}$$

Remark 2.2.3. The following are some additional remarks about Theorems 2.2.1 and 2.2.2.

- The excursion probability in (2.1.2) depends on the region where the maximum cross correlation is attained. In our setting, the maximum cross correlation ρ is attained on $D := \{(s,s) \mid s \in A_1 \cap A_2\}$.

- For Theorem 2.2.2, let us consider the extreme case when $M = 0$, i.e., $A_1 \cap A_2 = \{(T_1, ..., T_N)\}$. The exponential part still reaches $-\frac{u^2}{1+\rho}$, although the maximum cross correlation ρ is attained at a single point.

- To compare our results with [Ans06], we consider a centered Gaussian process $\{\mathbf{X}(t) = (X_1(t), X_2(t))^\top, t \in \mathbb{R}\}$ and $A_1 = A_2 = [0,T]$. In our setting, the cross correlation attains its maximum on the line $D = \{(s,s) \mid s \in [0,T]\}$, while in [Ans06] it only attains at a unique point in $[0,T] \times [0,T]$ because of the assumption **C2**. This is the reason why the power of u in our settings is $\frac{2}{\alpha_1} + \frac{2}{\alpha_2} - 3$ instead of $\frac{2}{\alpha_1} + \frac{2}{\alpha_2} - 4$ in [Ans06].

- Even though Theorem 2.2.2 only deals with a special case of A_1, A_2 with $mes_N(A_1 \cap A_2) = 0$, its method of proof can be applied to more general cases provided some information on A_1 and A_2 is provided. The key step is to reevaluate the infinite series in Lemma 2.4.5.

2.3 An example: positively correlated bivariate Matérn fields

In this section, we apply Theorems 2.2.1 and 2.2.2 to bivariate Gaussian random fields with the Matérn correlation functions introduced by [GKS10].

The Matérn correlation function $M(h|\nu,a)$, where $a>0,\nu>0$ are scale and smoothness parameters, is widely used to model covariance structures in spatial statistics. It is defined as

$$M(h|\nu,a) := \frac{2^{1-\nu}}{\Gamma(\nu)}(a|h|)^{\nu}K_{\nu}(a|h|), \tag{2.3.1}$$

where K_ν is a modified Bessel function of the second kind. In [GKS10], the authors introduce the full bivariate Matérn field $X(s)=(X_1(s),X_2(s))^\top$, i.e., an $\mathbb{R}^2$-valued Gaussian random field on $\mathbb{R}^N$ with zero mean and matrix-valued covariance functions:

$$C(h) = \begin{pmatrix} C_{11}(h) & C_{12}(h) \\ C_{21}(h) & C_{22}(h) \end{pmatrix}, \tag{2.3.2}$$

where $C_{ij}(h) := \mathbb{E}[X_i(s+h)X_j(s)]$ are specified by

$$\begin{aligned} C_{11}(h) &= \sigma_1^2 M(h|\nu_1,a_1), && (2.3.3)\\ C_{22}(h) &= \sigma_2^2 M(h|\nu_2,a_2), && (2.3.4)\\ C_{12}(h) &= C_{21}(h) = \rho\sigma_1\sigma_2 M(h|\nu_{12},a_{12}). && (2.3.5) \end{aligned}$$

According to [GKS10], the above model is valid if and only if

$$
\begin{aligned}
\rho^2 \leq & \frac{\Gamma(\nu_1+N/2)\Gamma(\nu_2+N/2)}{\Gamma(\nu_1)\Gamma(\nu_2)} \frac{\Gamma(\nu_{12})^2}{\Gamma(\nu_{12}+N/2)^2} \frac{a_1^{2\nu_1} a_2^{2\nu_2}}{a_{12}^{4\nu_{12}}} \\
& \times \inf_{t\geq 0} \frac{(a_{12}^2+t^2)^{2\nu_{12}+N}}{(a_1^2+t^2)^{\nu_1+N/2}(a_2^2+t^2)^{\nu_2+N/2}}.
\end{aligned}
\tag{2.3.6}
$$

Especially, when $a_1 = a_2 = a_{12}$, condition (2.3.6) is reduced to

$$
\rho^2 \leq \frac{\Gamma(\nu_1+N/2)\Gamma(\nu_2+N/2)}{\Gamma(\nu_1)\Gamma(\nu_2)} \frac{\Gamma(\nu_{12})^2}{\Gamma(\nu_{12}+N/2)^2}, \tag{2.3.7}
$$

in which case the choice of ρ is fairly flexible.

Here we focus on a standardized bivariate Matérn field, that is, we assume $\sigma_1 = \sigma_2 = 1$, $a_1 = a_2 = a_{12} = 1$ and $\rho > 0$. Moreover, we assume $\nu_1, \nu_2 \in (0, 1)$ and $\nu_{12} > 1$. In this case, the bivariate Matérn field $\{\mathbf{X}(t),\ t \in \mathbb{R}^N\}$ satisfies the assumptions in Section 2.1.

Indeed, Assumption i) in Section 2.1 is satisfied since

$$
M(h|\nu_i, a) = 1 - c_i|t|^{2\nu_i} + o(|t|^{2\nu_i}),
$$

where $c_i = \frac{\Gamma(1-\nu_i)}{2^{2\nu_i}\Gamma(1+\nu_i)}$ (see, e.g., [Ste99], p. 32). Assumption ii) holds immediately if we use the following integral representation of $M(h|\nu, a)$ (see, e.g., [AS72], Section 9.6)

$$
M(h|\nu, a) = \frac{2\Gamma(\nu+1/2)}{\sqrt{\pi}\Gamma(\nu)} \int_0^\infty \frac{\cos(a|h|r)}{(1+r^2)^{\nu+1/2}}\, dr. \tag{2.3.8}
$$

Assumption iii) holds by the definition of cross correlation in (2.3.5). For Assumption iv), we only need to check the smoothness of $M(h|\nu, a)$. By another integral representation of

$M(h|\nu, a)$ (see, e.g., [AS72], Section 9.6), i.e.,

$$M(h|\nu, a) = \frac{2^{1-2\nu}(a|h|)^{2\nu}}{\Gamma(\nu+1/2)\Gamma(\nu)} \int_1^\infty e^{-a|h|r}(r^2-1)^{\nu-1/2}\, dr,$$

one can verify that $M(h|\nu, a)$ is infinitely differentiable when $|h| \neq 0$. Meanwhile, $M''(0|\nu, a)$ exists and is continuous when $\nu > 1$ which can be proven by taking twice derivatives to the integral representation in (2.3.8) w.r.t. $|h|$. So Assumption iv) holds.

Applying Theorem 2.2.1 to the double excursion probability of $X(s)$ over $[0,1]^N$, we have

$$\begin{aligned}
&\mathbb{P}\Big(\max_{s\in[0,1]^N} X_1(s) > u, \max_{t\in[0,1]^N} X_2(t) > u\Big) \\
&= (2\pi)^{\frac{N}{2}}(-C_{12}''(0))^{-\frac{N}{2}} c_1^{\frac{N}{2\nu_1}} c_2^{\frac{N}{2\nu_2}} (1+\rho)^{-N(\frac{1}{\nu_1}+\frac{1}{\nu_2}-1)} H_{2\nu_1}H_{2\nu_2} \\
&\quad \times u^{N(\frac{1}{\nu_1}+\frac{1}{\nu_2}-1)}\Psi(u,\rho)(1+o(1)), \quad \text{as } u\to\infty.
\end{aligned}$$

Secondly, when the two measurements are observed on two regions which only share part of boundaries, we use Theorem 2.2.2 to obtain the excursion probability. For example, if $X_1(s)$ are observed on the region $[0,1]^N$ and $X_2(s)$ on $[0,1]^{N-1}\times[1,2]$, then as $u\to\infty$,

$$\begin{aligned}
&\mathbb{P}\bigg(\max_{s\in[0,1]^N} X_1(s) > u, \max_{t\in[0,1]^{N-1}\times[1,2]} X_2(t) > u\bigg) \\
&= (2\pi)^{\frac{N-1}{2}}(-C_{12}''(0))^{-\frac{N+1}{2}} c_1^{\frac{N}{2\nu_1}} c_2^{\frac{N}{2\nu_2}} (1+\rho)^{1-N(\frac{1}{\nu_1}+\frac{1}{\nu_2}-1)} H_{2\nu_1}H_{2\nu_2} \\
&\quad \times u^{N(\frac{1}{\nu_1}+\frac{1}{\nu_2}-1)-1}\Psi(u,\rho)(1+o(1)).
\end{aligned}$$

2.4 Proofs of the main results

The proofs of Theorems 2.2.1 and 2.2.2 are based on the double sum method [Pit96] and the work of [LP00]. Since the latter deals with the tail probability $\mathbb{P}(\max_{t\in[T_1,T_2]} X(t) > u, \max_{t\in[T_3,T_4]} X(t) > u)$ of a univariate Gaussian process $\{X(t), t \in \mathbb{R}\}$, their method is not sufficient for carrying out the double sum method for a bivariate random field.

The lemmas below extend Lemma 1 and Lemma 9 in [LP00] to the bivariate random field $\{(X_1(t), X_2(t))^\top, t \in \mathbb{R}^N\}$. Moreover, we have strengthened the conclusion by showing that the convergence is uniform in certain sense. This will be useful later for dealing with sums of local approximations around the regions where the maximum cross correlation is attained. The details will be illustrated in the proof of Theorem 2.2.1 (see, e.g., (2.4.10), (2.4.21)). In the following lemmas, $\{\mathbf{X}(t), t \in \mathbb{R}^N\}$ is a bivariate Gaussian random field as defined in Section 2.1.

Lemma 2.4.1. *Let s_u and t_u be two $\mathbb{R}^N$-valued functions of u and let $\tau_u := t_u - s_u$. For any compact sets $\mathbb{S}$ and $\mathbb{T}$ in $\mathbb{R}^N$, we have*

$$\begin{aligned}
&\mathbb{P}\left(\max_{s\in s_u+u^{-2/\alpha_1}\mathbb{S}} X_1(s) > u, \max_{t\in t_u+u^{-2/\alpha_2}\mathbb{T}} X_2(t) > u\right)\\
&= \frac{(1+\rho)^2}{2\pi\sqrt{1-\rho^2}} H_{\alpha_1}\left(\frac{c_1^{1/\alpha_1}\mathbb{S}}{(1+\rho)^{\frac{2}{\alpha_1}}}\right) H_{\alpha_2}\left(\frac{c_2^{1/\alpha_2}\mathbb{T}}{(1+\rho)^{\frac{2}{\alpha_2}}}\right)\\
&\quad\times u^{-2}\exp\left(-\frac{u^2}{1+r(|\tau_u|)}\right)(1+o(1)),
\end{aligned} \tag{2.4.1}$$

where $o(1) \to 0$ uniformly w.r.t. τ_u satisfying $|\tau_u| \le C\sqrt{\log u}/u$ as $u \to \infty$.

Lemma 2.4.2. *Let s_u, t_u and τ_u be the same as in Lemma 2.4.1. For all $T > 0$, $\mathbf{m}, \mathbf{n} \in \mathbb{Z}^N$,* *we have*

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{s \in s_u + u^{-2/\alpha_1}[0,T]^N} X_1(s) > u, \max_{t \in t_u + u^{-2/\alpha_2}[0,T]^N} X_2(t) > u, \\
&\qquad \max_{s \in s_u + u^{-2/\alpha_1}[\mathbf{m}T, (\mathbf{m}+1)T]} X_1(s) > u, \max_{t \in t_u + u^{-2/\alpha_2}[\mathbf{n}T, (\mathbf{n}+1)T]} X_2(t) > u \Big) \\
&= \frac{(1+\rho)^2}{2\pi\sqrt{1-\rho^2}\, u^2}\, e^{-\frac{u^2}{1+r(|\tau_u|)}} H_{\alpha_1}\bigg(\frac{c_1^{1/\alpha_1}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_1}}}, \frac{c_1^{1/\alpha_1}[\mathbf{m}T, (\mathbf{m}+1)T]}{(1+\rho)^{\frac{2}{\alpha_1}}} \bigg) \\
&\qquad \times H_{\alpha_2}\bigg(\frac{c_2^{1/\alpha_2}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_2}}}, \frac{c_2^{1/\alpha_2}[\mathbf{n}T, (\mathbf{n}+1)T]}{(1+\rho)^{\frac{2}{\alpha_2}}} \bigg) \big(1+o(1)\big), \qquad (2.4.2)
\end{aligned}
$$

where $H_\alpha(\cdot,\cdot)$ *is defined in* (2.2.1) *and* $o(1) \to 0$ *uniformly for all* s_u *and* t_u *that satisfy* $|\tau_u| \le C\sqrt{\log u}/u$ *as* $u \to \infty$.

Now we are ready to prove our main theorems.

Proof of Theorem 2.2.1. Let $\Pi = A_1 \times A_2$, $\delta(u) = C\sqrt{\log u}/u$, where C is a constant whose value will be determined later. Let

$$
\mathcal{D} = \big\{ (s,t) \in \Pi : |t-s| \le \delta(u) \big\}. \qquad (2.4.3)
$$

Since

$$
\begin{aligned}
&\mathbb{P}\Big(\bigcup_{(s,t)\in\mathcal{D}} \{X_1(s) > u, X_2(t) > u)\} \Big) \le \mathbb{P}\Big(\max_{s\in A_1} X_1(s) > u, \max_{t \in A_2} X_2(t) > u \Big) \\
&\le \mathbb{P}\Big(\bigcup_{(s,t)\in\mathcal{D}} \{X_1(s) > u, X_2(t) > u)\} \Big) + \mathbb{P}\Big(\bigcup_{(s,t)\in\Pi\setminus\mathcal{D}} \{X_1(s) > u, X_2(t) > u)\} \Big),
\end{aligned}
$$

it is sufficient to prove that, by choosing appropriate constant C, we have

$$\begin{aligned}
&\mathbb{P}\Big(\bigcup_{(s,t)\in\mathcal{D}}\{X_1(s)>u, X_2(t)>u)\}\Big)\\
&=(2\pi)^{\frac{N}{2}}(-r''(0))^{-\frac{N}{2}}c_1^{\frac{N}{\alpha_1}}c_2^{\frac{N}{\alpha_2}}(1+\rho)^{-N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}mes_N(A_1\cap A_2)\\
&\quad\times H_{\alpha_1}H_{\alpha_2}u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)(1+o(1)),\quad \text{as } u\to\infty
\end{aligned}\tag{2.4.4}$$

and

$$\lim_{u\to\infty}\frac{\mathbb{P}\left(\bigcup_{(s,t)\in\Pi\setminus\mathcal{D}}\{X_1(s)>u, X_2(t)>u)\}\right)}{\mathbb{P}\left(\bigcup_{(s,t)\in\mathcal{D}}\{X_1(s)>u, X_2(t)>u)\}\right)}=0. \tag{2.4.5}$$

We prove (2.4.4) first. For any fixed $T>0$ and $i=1,2$, let $d_i(u)=Tu^{-\frac{2}{\alpha_i}}$ and, for any $\mathbf{k}=(k_1,\dots,k_N)\in\mathbb{Z}^N$, define

$$\Delta_{\mathbf{k}}^{(i)}\triangleq\prod_{j=1}^{N}[k_jd_i(u),(k_j+1)d_i(u)]=[\mathbf{k}d_i(u),(\mathbf{k}+1)d_i(u)]. \tag{2.4.6}$$

Let

$$\mathcal{C}=\{(\mathbf{k},\mathbf{l}):\Delta_{\mathbf{k}}^{(1)}\times\Delta_{\mathbf{l}}^{(2)}\cap\mathcal{D}\neq\emptyset\}\ \text{ and }\ \mathcal{C}^\circ=\{(\mathbf{k},\mathbf{l}):\Delta_{\mathbf{k}}^{(1)}\times\Delta_{\mathbf{l}}^{(2)}\subseteq\mathcal{D}\}. \tag{2.4.7}$$

It is easy to see that

$$\bigcup_{(\mathbf{k},\mathbf{l})\in\mathcal{C}^\circ}\Delta_{\mathbf{k}}^{(1)}\times\Delta_{\mathbf{l}}^{(2)}\subseteq\mathcal{D}\subseteq\bigcup_{(\mathbf{k},\mathbf{l})\in\mathcal{C}}\Delta_{\mathbf{k}}^{(1)}\times\Delta_{\mathbf{l}}^{(2)}.$$

Thus the LHS of (2.4.4) is bounded above by

$$
\begin{aligned}
&\mathbb{P}\Big(\bigcup_{(s,t)\in\mathcal{D}}\{X_1(s)>u, X_2(t)>u)\}\Big)\\
&\le \sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}}\mathbb{P}\Big(\max_{s\in\Delta_{\mathbf{k}}^{(1)}}X_1(s)>u,\ \max_{t\in\Delta_{\mathbf{l}}^{(2)}}X_2(t)>u\Big)\\
&=\sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}}\mathbb{P}\Big(\max_{s\in\mathbf{k}d_1(u)+\Delta_0^{(1)}}X_1(s)>u,\ \max_{t\in\mathbf{l}d_2(u)+\Delta_0^{(2)}}X_2(t)>u\Big).
\end{aligned}
\tag{2.4.8}
$$

Let

$$
\begin{aligned}
\tau_{\mathbf{kl}} &:= \mathbf{l}d_2(u)-\mathbf{k}d_1(u)\\
&=(l_1d_2(u)-k_1d_1(u),...,l_Nd_2(u)-k_Nd_1(u)).
\end{aligned}
\tag{2.4.9}
$$

For $(\mathbf{k},\mathbf{l})\in\mathcal{C}$, $|\tau_{\mathbf{kl}}|\le\delta(u)+\sqrt{N}(d_1(u)+d_2(u))\le 2\delta(u)$ for all u large enough, since $d_1(u)=o(\delta(u))$ and $d_2(u)=o(\delta(u))$, as $u\to\infty$. Hence, by applying Lemma 2.4.1 to the RHS of (2.4.8), we obtain

$$
\begin{aligned}
&\mathbb{P}\Big(\bigcup_{(s,t)\in\mathcal{D}}\{X_1(s)>u, X_2(t)>u)\}\Big)\\
&\le \frac{(1+\rho)^2(1+\gamma(u))}{2\pi\sqrt{1-\rho^2}\,u^2}H_{\alpha_1}\left(\frac{c_1^{1/\alpha_1}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_1}}}\right)H_{\alpha_2}\left(\frac{c_2^{1/\alpha_2}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_2}}}\right)\\
&\qquad\times\sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}}\exp\left(-\frac{u^2}{1+r(|\tau_{\mathbf{kl}}|)}\right)\\
&=H_{\alpha_1}\left(\frac{c_1^{1/\alpha_1}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_1}}}\right)H_{\alpha_2}\left(\frac{c_2^{1/\alpha_2}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_2}}}\right)\Psi(u,\rho)(1+\gamma(u))\\
&\qquad\times\sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}}\exp\left\{-u^2\left(\frac{1}{1+r(|\tau_{\mathbf{kl}}|)}-\frac{1}{1+\rho}\right)\right\},
\end{aligned}
\tag{2.4.10}
$$

where the global error function $\gamma(u) \to 0$, as $u \to \infty$. The uniform convergence of (2.4.1) in Lemma 2.4.1 guarantees that the local error term $o(1)$ for each pair $(\mathbf{k}, \mathbf{l}) \in \mathcal{C}$ is uniformly bounded by $\gamma(u)$.

The series in the last equality of (2.4.10) is dealt by the following key lemma, which gives the power of the threshold u in (2.4.4).

Lemma 2.4.3. *Recall the set* $\mathcal{C}$ *defined in* (2.4.7). *Let*

$$h(u) := \sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}} \exp\left\{-u^2\left(\frac{1}{1+r(|\tau_{\mathbf{kl}}|)} - \frac{1}{1+\rho}\right)\right\}. \tag{2.4.11}$$

Then, under the assumptions of Theorem 2.2.1, we have

$$\begin{aligned} h(u) = &(2\pi)^{N/2}(-r''(0))^{-N/2}(1+\rho)^N T^{-2N} mes_N(A_1 \cap A_2) \\ &\times u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}(1+o(1)), \quad as \ \ u \to \infty. \end{aligned} \tag{2.4.12}$$

Moreover, if we replace $\mathcal{C}$ *in (2.4.11) by* $\mathcal{C}^\circ$ *defined in* (2.4.7), *then (2.4.12) still holds.*

We defer the proof of Lemma 2.4.3 to Section 2.5 and continue with the proof of Theorem 2.2.1. Applying (2.4.12) to (2.4.10), we obtain

$$\begin{aligned} &\mathbb{P}\Bigg(\bigcup_{(s,t)\in\mathcal{D}}\{X_1(s) > u, X_2(t) > u)\}\Bigg) \\ &\le (2\pi)^{\frac{N}{2}}(-r''(0))^{-\frac{N}{2}}(1+\rho)^N T^{-2N} mes_N(A_1\cap A_2) H_{\alpha_1}\left(\frac{c_1^{1/\alpha_1}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_1}}}\right) \\ &\quad \times H_{\alpha_2}\left(\frac{c_2^{1/\alpha_2}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_2}}}\right) u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)(1+\gamma_1(u)), \end{aligned} \tag{2.4.13}$$

where $\gamma_1(u) \to 0$, as $u \to \infty$. Hence,

$$\begin{aligned}
&\limsup_{u\to\infty} \frac{\mathbb{P}\left(\bigcup_{(s,t)\in\mathcal{D}}\{X_1(s) > u, X_2(t) > u)\}\right)}{u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)} \\
&\le (2\pi)^{\frac{N}{2}}(-r''(0))^{-\frac{N}{2}}(1+\rho)^N mes_N(A_1\cap A_2) \qquad (2.4.14)\\
&\quad \times T^{-2N} H_{\alpha_1}\left(\frac{c_1^{1/\alpha_1}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_1}}}\right) H_{\alpha_2}\left(\frac{c_2^{1/\alpha_2}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_2}}}\right).
\end{aligned}$$

The above inequality holds for every $T > 0$. Therefore, letting $T \to \infty$, we have

$$\begin{aligned}
&\limsup_{u\to\infty} \frac{\mathbb{P}\left(\bigcup_{(s,t)\in\mathcal{D}}\{X_1(s) > u, X_2(t) > u)\}\right)}{u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)} \le (2\pi)^{\frac{N}{2}}(-r''(0))^{-\frac{N}{2}} \\
&\quad \times c_1^{\frac{N}{\alpha_1}} c_2^{\frac{N}{\alpha_2}}(1+\rho)^{-N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)} mes_N(A_1\cap A_2) H_{\alpha_1} H_{\alpha_2}. \qquad (2.4.15)
\end{aligned}$$

On the other hand, the lower bound for LHS of (2.4.4) can be derived as follows. Let

$$\mathcal{B} = \{(\mathbf{k},\mathbf{l},\mathbf{k}',\mathbf{l}') :\ (\mathbf{k},\mathbf{l}) \neq (\mathbf{k}',\mathbf{l}'), (\mathbf{k},\mathbf{l}), (\mathbf{k}',\mathbf{l}') \in \mathcal{C}\}. \qquad (2.4.16)$$

By Bonferroni's inequality and symmetric property of $\mathcal{B}$, the LHS of (2.4.4) is bounded below by

$$\begin{aligned}
&\mathbb{P}\Big(\bigcup_{(s,t)\in\mathcal{D}}\{X_1(s) > u, X_2(t) > u\}\Big) \\
&\ge \sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}^\circ} \mathbb{P}\Big(\max_{s\in\Delta_{\mathbf{k}}^{(1)}} X_1(s) > u,\ \max_{t\in\Delta_{\mathbf{l}}^{(2)}} X_2(t) > u\Big) \\
&\quad - \frac{1}{2}\sum_{(\mathbf{k},\mathbf{l},\mathbf{k}',\mathbf{l}')\in\mathcal{B}} \mathbb{P}\Big(\max_{s\in\Delta_{\mathbf{k}}^{(1)}} X_1(s) > u,\ \max_{t\in\Delta_{\mathbf{l}}^{(2)}} X_2(t) > u, \qquad (2.4.17)
\end{aligned}$$

$$\max_{s\in\Delta^{(1)}_{\mathbf{k}'}} X_1(s) > u, \ \max_{t\in\Delta^{(2)}_{\mathbf{l}'}} X_2(t) > u\Big)$$

$$\triangleq \Sigma_1 - \Sigma_2.$$

Since $\mathcal{C}^\circ$ and $\mathcal{C}$ are almost the same, a similar argument as in (2.4.10)$\sim$ (2.4.15) shows that Σ_1 is bounded from below by

$$\Sigma_1 \geq (2\pi)^{\frac{N}{2}}(-r''(0))^{-\frac{N}{2}}(1+\rho)^N mes_N(A_1\cap A_2)T^{-2N} H_{\alpha_1}\left(\frac{c_1^{1/\alpha_1}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_1}}}\right)$$
$$\times H_{\alpha_2}\left(\frac{c_2^{1/\alpha_2}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha_2}}}\right) u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)(1-\gamma_2(u)), \tag{2.4.18}$$

where $\gamma_2(u)\to 0$, as $u\to\infty$. Hence, letting $T\to\infty$, we have

$$\begin{aligned}\liminf_{u\to\infty}\frac{\Sigma_1}{u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)} &\geq (2\pi)^{\frac{N}{2}}(-r''(0))^{-\frac{N}{2}}c_1^{\frac{N}{\alpha_1}}c_2^{\frac{N}{\alpha_2}} \\ &\times (1+\rho)^{-N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)} mes_N(A_1\cap A_2)H_{\alpha_1}H_{\alpha_2}.\end{aligned} \tag{2.4.19}$$

Next, we consider Σ_2 in (2.4.17). To simplify the notation, we let

$$I(\mathbf{k},\mathbf{l},\mathbf{k}',\mathbf{l}') := \mathbb{P}\Big(\max_{s\in\Delta^{(1)}_{\mathbf{k}}} X_1(s) > u, \ \max_{t\in\Delta^{(2)}_{\mathbf{l}}} X_2(t) > u,$$
$$\max_{s\in\Delta^{(1)}_{\mathbf{k}'}} X_1(s) > u, \ \max_{t\in\Delta^{(2)}_{\mathbf{l}'}} X_2(t) > u\Big).$$

For $\mathbf{m}=(m_1,\dots,m_N)\in\mathbb{Z}^N$, let

$$\mathcal{H}_{\alpha,c}(\mathbf{m}) \triangleq H_\alpha\left(\frac{c^{1/\alpha}[0,T]^N}{(1+\rho)^{\frac{2}{\alpha}}}, \frac{c^{1/\alpha}[\mathbf{m}T,(\mathbf{m}+1)T]}{(1+\rho)^{\frac{2}{\alpha}}}\right). \tag{2.4.20}$$

Rewriting Σ_2 and applying Lemma 2.4.2, we obtain

$$
\begin{aligned}
\Sigma_2 = &\frac{1}{2} \sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}} \Bigg(\sum_{\substack{(\mathbf{k}',\mathbf{l}')\in\mathcal{C} \\ \mathbf{k}'=\mathbf{k},\mathbf{l}'\neq\mathbf{l}}} + \sum_{\substack{(\mathbf{k}',\mathbf{l}')\in\mathcal{C} \\ \mathbf{k}'\neq\mathbf{k},\mathbf{l}'=\mathbf{l}}} + \sum_{\substack{(\mathbf{k}',\mathbf{l}')\in\mathcal{C} \\ \mathbf{k}'\neq\mathbf{k},\mathbf{l}'\neq\mathbf{l}}} \Bigg) I(\mathbf{k},\mathbf{l},\mathbf{k}',\mathbf{l}') \\
= &\frac{(1+\rho)^2(1+\gamma_3(u))}{4\pi\sqrt{1-\rho^2}\,u^2} \sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}} e^{-\frac{u^2}{1+r(|\tau_{\mathbf{kl}}|)}} \Bigg(\mathcal{H}_{\alpha_1,c_1}(\mathbf{0}) \sum_{\substack{(\mathbf{k}',\mathbf{l}')\in\mathcal{C} \\ \mathbf{k}'=\mathbf{k},\mathbf{l}'\neq\mathbf{l}}} \mathcal{H}_{\alpha_2,c_2}(\mathbf{l}'-\mathbf{l}) \\
&+ \mathcal{H}_{\alpha_2,c_2}(\mathbf{0}) \sum_{\substack{(\mathbf{k}',\mathbf{l}')\in\mathcal{C} \\ \mathbf{k}'\neq\mathbf{k},\mathbf{l}'=\mathbf{l}}} \mathcal{H}_{\alpha_1,c_1}(\mathbf{k}'-\mathbf{k}) + \sum_{\substack{(\mathbf{k}',\mathbf{l}')\in\mathcal{C} \\ \mathbf{k}'\neq\mathbf{k},\mathbf{l}'\neq\mathbf{l}}} \mathcal{H}_{\alpha_1,c_1}(\mathbf{k}'-\mathbf{k})\mathcal{H}_{\alpha_2,c_2}(\mathbf{l}'-\mathbf{l}) \Bigg) \\
\leq &\frac{(1+\rho)^2(1+\gamma_3(u))}{4\pi\sqrt{1-\rho^2}u^2} \sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}} e^{-\frac{u^2}{1+r(|\tau_{\mathbf{kl}}|)}} \Bigg(\mathcal{H}_{\alpha_1,c_1}(\mathbf{0}) \sum_{\mathbf{n}\neq\mathbf{0}} \mathcal{H}_{\alpha_2,c_2}(\mathbf{n}) \\
&+ \mathcal{H}_{\alpha_2,c_2}(\mathbf{0}) \sum_{\mathbf{m}\neq\mathbf{0}} \mathcal{H}_{\alpha_1,c_1}(\mathbf{m}) + \sum_{\mathbf{m}\neq\mathbf{0},\mathbf{n}\neq\mathbf{0}} \mathcal{H}_{\alpha_1,c_1}(\mathbf{m})\mathcal{H}_{\alpha_2,c_2}(\mathbf{n}) \Bigg),
\end{aligned}
\tag{2.4.21}
$$

where $\gamma_3(u) \to 0$, as $u \to \infty$. According to the uniform convergence of (2.4.2), the local error term $o(1)$ for each pair $(\mathbf{k}',\mathbf{l}') \in \mathcal{C}$ is bounded above by $\gamma_3(u)$. To estimate $\mathcal{H}_{\alpha,c}(\cdot)$, we make use of the following lemma, whose proof is again postponed to Section 2.5.

Lemma 2.4.4. *Recall $\mathcal{H}_{\alpha,c}(\cdot)$ defined in* (2.4.20)*. Let $i_0 = \mathrm{argmax}_{1\leq i\leq N}|m_i|$. Then there exist positive constants C_1 and T_0 such that for all $T \geq T_0$,*

$$\mathcal{H}_{\alpha,c}(\mathbf{0}) \leq C_1 T^N; \tag{2.4.22}$$

$$\mathcal{H}_{\alpha,c}(\mathbf{m}) \leq C_1 T^{N-\frac{1}{2}}, \text{ when } |m_{i_0}| = 1; \tag{2.4.23}$$

$$\mathcal{H}_{\alpha,c}(\mathbf{m}) \leq C_1 T^{2N} e^{-\frac{c}{8(1+\rho)^2}(|m_{i_0}|-1)^\alpha T^\alpha}, \text{ when } |m_{i_0}| \geq 2. \tag{2.4.24}$$

Consequently,

$$\sum_{\mathbf{m}\in\mathbb{Z}^N\setminus\{\mathbf{0}\}} \mathcal{H}_{\alpha,c}(\mathbf{m}) \le C_1 T^{N-\frac{1}{2}}. \tag{2.4.25}$$

Applying Lemmas 2.4.3 and 2.4.4 to the RHS of (2.4.21), we obtain

$$\begin{aligned}\Sigma_2 &\le \frac{C_0(1+\rho)^2(1+\gamma_3(u))}{4\pi\sqrt{1-\rho^2}u^2} T^{2N-\frac{1}{2}} \sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}} \exp\left(-\frac{u^2}{1+r(|\tau_{\mathbf{kl}}|)}\right)\\ &\le C_0(2\pi)^{\frac{N}{2}}(-r''(0))^{-\frac{N}{2}}(1+\rho)^N mes_N(A_1\cap A_2)T^{-\frac{1}{2}}\\ &\quad \times u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)(1+\gamma_4(u)),\end{aligned} \tag{2.4.26}$$

where $\gamma_4(u)\to 0$, as $u\to\infty$. By letting $u\to\infty$ and $T\to\infty$ successively, we have

$$\limsup_{u\to\infty} \frac{\Sigma_2}{u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)} = 0. \tag{2.4.27}$$

By combining (2.4.17), (2.4.19) and (2.4.27), we have

$$\begin{aligned}&\liminf_{u\to\infty} \frac{\mathbb{P}\left(\bigcup_{(s,t)\in\mathcal{D}}\{X_1(s)>u, X_2(t)>u)\}\right)}{u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)}\\ &\ge \liminf_{u\to\infty}\frac{\Sigma_1}{u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)} - \limsup_{u\to\infty}\frac{\Sigma_2}{u^{N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)}\Psi(u,\rho)}\\ &\ge (2\pi)^{\frac{N}{2}}(-r''(0))^{-\frac{N}{2}} c_1^{\frac{N}{\alpha_1}} c_2^{\frac{N}{\alpha_2}}(1+\rho)^{-N(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1)} mes_N(A_1\cap A_2) H_{\alpha_1}H_{\alpha_2}.\end{aligned} \tag{2.4.28}$$

It is now clear that (2.4.4) follows from (2.4.15) and (2.4.28).

Now we prove (2.4.5). Define

$$Y(s,t) := X_1(s) + X_2(t), \text{ for } (s,t)\in\Pi\setminus\mathcal{D}. \tag{2.4.29}$$

For $x = (s_1, t_1), y = (s_2, t_2) \in \Pi \setminus \mathcal{D}$, let $|x - y| = \sqrt{|s_1 - s_2|^2 + |t_1 - t_2|^2}$. Then we can verify that

$$\mathbb{E}|Y(x) - Y(y)|^2 \leq C_0 |x - y|^{\min(\alpha_1, \alpha_2)}, \ \forall x, y \in \Pi \setminus \mathcal{D}. \tag{2.4.30}$$

By applying Theorem 8.1 in [Pit96], we obtain that the numerator of (2.4.5) is bounded above by

$$\begin{aligned} &\mathbb{P}\Bigg(\bigcup_{(s,t) \in \Pi \setminus \mathcal{D}} \{X_1(s) > u, X_2(t) > u)\} \Bigg) \leq \mathbb{P}\left(\max_{(s,t) \in \Pi \setminus \mathcal{D}} Y(s,t) > 2u \right) \\ &\leq C_0 u^{-1 + \frac{2N}{\min(\alpha_1, \alpha_2)}} \exp\left(- \frac{u^2}{1 + \max_{(s,t) \in \Pi \setminus \mathcal{D}} r(|t - s|)} \right). \end{aligned} \tag{2.4.31}$$

Since $r(|t - s|) = \rho + \frac{1}{2} r''(0) |t - s|^2 (1 + o(1))$ and $r(\cdot)$ attains maximum only at zero, we have

$$\max_{(s,t) \in \Pi \setminus \mathcal{D}} r(|t - s|) \leq \rho - \frac{1}{3}(-r''(0)) \delta^2(u) \tag{2.4.32}$$

for u large enough. So (2.4.31) is at most

$$\begin{aligned} &C_0 u^{-1 + \frac{2N}{\min(\alpha_1, \alpha_2)}} \exp\left(- \frac{u^2}{1 + \rho - \frac{1}{3}(-r''(0)) \delta^2(u)} \right) \\ &\leq C_0 u^{-1 + \frac{2N}{\min(\alpha_1, \alpha_2)}} \exp\left(- \frac{u^2}{1 + \rho} \right) \exp\left(- \frac{\frac{1}{3}(-r''(0)) \delta^2(u) u^2}{(1 + \rho)^2} \right) \\ &= \frac{2\pi \sqrt{1 - \rho^2} C_0}{(1 + \rho)^2} u^{1 + \frac{2N}{\min(\alpha_1, \alpha_2)} - \frac{-r''(0)}{3(1+\rho)^2} C^2} \Psi(u, \rho), \end{aligned} \tag{2.4.33}$$

where the inequality holds since $\frac{1}{x-y} \geq \frac{1}{x} + \frac{y}{x^2}, \forall x > y$. Compare (2.4.33) with (2.4.4), it is easy to see (2.4.5) holds if and only if

$$1+\frac{2N}{\min(\alpha_1,\alpha_2)}-\frac{-r''(0)}{3(1+\rho)^2}C^2<N\Big(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-1\Big) \tag{2.4.34}$$

Hence, by choosing the constant C satisfying

$$C>\left[\frac{3(1+\rho)^2}{-r''(0)}\left(N\Big(\frac{2}{\min(\alpha_1,\alpha_2)}+1-\frac{2}{\alpha_1}-\frac{2}{\alpha_2}\Big)+1\right)_+\right]^{\frac{1}{2}}, \tag{2.4.35}$$

we conclude (2.4.5). □

Proof of Theorem 2.2.2. From the proof of Theorem 2.2.1, we see that the exponential decaying rate of the excursion probability is only determined by the region where the maximum cross correlation is attained. In the case of $mes_N(A_1\cap A_2)=0$ but $A_1\cap A_2\neq\emptyset$, the exponential part, $e^{-\frac{u^2}{1+\rho}}$, remains the same. Yet, the dimension reduction of $A_1\cap A_2$ does affect the polynomial power of the excursion probability, which is determined by the quantity

$$h(u)=\sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}}\exp\left\{-u^2\left(\frac{1}{1+r(|\tau_{\mathbf{kl}}|)}-\frac{1}{1+\rho}\right)\right\}$$

in Lemma 2.4.3. Under the assumptions of Theorem 2.2.2, the set $\mathcal{C}$ and the behavior of $h(u)$ change. We will make use of the following lemma which plays the role of Lemma 2.4.3.

Lemma 2.4.5. *Under the assumptions of Theorem 2.2.2, we have*

$$\begin{aligned}h(u)&=(2\pi)^{M/2}(-r''(0))^{M/2-N}(1+\rho)^{2N-M}T^{-2N}mes_M(A_{1,M}\cap A_{2,M})\\&\quad\times u^{M+N\left(\frac{2}{\alpha_1}+\frac{2}{\alpha_2}-2\right)}(1+o(1)),\ \text{as } u\to\infty.\end{aligned} \tag{2.4.36}$$

Moreover, if we replace $\mathcal{C}$ with $\mathcal{C}^\circ$ defined in (2.4.7)*, then the above statement still holds.*

The rest of the proof of Theorem 2.2.2 is the same as that of Theorem 2.2.1 and it is omitted here. □

2.5 Proof of Lemmas

For proving Lemma 2.4.1, we will make use of the following

Lemma 2.5.1. *Let s_u and t_u be two $\mathbb{R}^N$-valued functions of u and let $\tau_u := t_u - s_u$. For any compact rectangles $\mathbb{S}$ and $\mathbb{T}$ in $\mathbb{R}^N$, define*

$$\begin{aligned}
\xi_u(s) &:= u(X_1(s_u + u^{-2/\alpha_1}s) - u) + x, \quad \forall\, s \in \mathbb{S}, \\
\eta_u(t) &:= u(X_2(t_u + u^{-2/\alpha_2}t) - u) + y, \quad \forall\, t \in \mathbb{T}
\end{aligned} \tag{2.5.1}$$

and for any $t \in \mathbb{R}^N$, let

$$\xi(t) := \sqrt{c_1}\chi_1(t) - \frac{c_1|t|^{\alpha_1}}{1+\rho}, \quad \eta(t) := \sqrt{c_2}\chi_2(t) - \frac{c_2|t|^{\alpha_2}}{1+\rho}, \tag{2.5.2}$$

where $\chi_1(t), \chi_2(t)$ are two independent fractional Brownian motions with indices $\alpha_1/2$ and $\alpha_2/2$, respectively. Then, the finite dimensional distributions (abbr. f.d.d.) of $(\xi_u(\cdot), \eta_u(\cdot))$, given $X_1(s_u) = u - \frac{x}{u}$, $X_2(t_u) = u - \frac{y}{u}$, converge uniformly to the f.d.d. of $(\xi(\cdot), \eta(\cdot))$ for all s_u and t_u that satisfy $|\tau_u| \le C\sqrt{\log u}/u$. Furthermore, as $u \to \infty$,

$$\begin{aligned}
&\mathbb{P}\bigg(\max_{s\in\mathbb{S}} \xi_u(s) > x, \max_{t\in\mathbb{T}} \eta_u(t) > y \;\bigg|\; X_1(s_u) = u - \frac{x}{u}, X_2(t_u) = u - \frac{y}{u} \bigg) \\
&\to \mathbb{P}\bigg(\max_{s\in\mathbb{S}} \xi(s) > x, \max_{t\in\mathbb{T}} \eta(t) > y \bigg),
\end{aligned} \tag{2.5.3}$$

where the convergence is uniform for all s_u and t_u that satisfy $|\tau_u| \le C\sqrt{\log u}/u$.

Proof. First, we prove the uniform convergence of finite dimensional distributions. Given $X_1(s_u) = u - \frac{x}{u}$, $X_2(t_u) = u - \frac{y}{u}$, the distribution of the bivariate random field $(\xi_u(\cdot), \eta_u(\cdot))$ is still Gaussian. Thanks to the following lemma (whose proof will be given at the end of this section), it suffices to prove the uniform convergence of conditional mean and conditional variance.

Lemma 2.5.2. *Let $X(u,\tau_u) = (X_1(u,\tau_u), \ldots, X_n(u,\tau_u))^\top$ be a Gaussian random vector with mean $\mu(u,\tau_u) = (\mu_1(u,\tau_u), ..., \mu_n(u,\tau_u))^\top$ and covariance matrix $\Sigma(u,\tau_u)$ with entries $\sigma_{ij}(u,\tau_u) = \mathrm{Cov}(X_i(u,\ \tau_u), X_j(u,\tau_u))$, $i,j = 1,2,\ldots,n$. Similarly, let $X = (X_1,\ldots,X_n)^\top$ be a Gaussian random vector with mean $\mu = (\mu_1, ..., \mu_n)$ and covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1}^n$. Let $F_u(\cdot)$ and $F(\cdot)$ be the distribution functions of $X(u,\tau_u)$ and X respectively. If*

$$
\begin{aligned}
&\lim_{u\to\infty} \max_{\tau_u} |\mu_j(u,\tau_u) - \mu_j| = 0,\\
&\lim_{u\to\infty} \max_{\tau_u} |\sigma_{ij}(u,\tau_u) - \sigma_{ij}| = 0, \quad i,j = 1,2,\ldots,n,
\end{aligned} \tag{2.5.4}
$$

then for any $x \in \mathbb{R}^N$,

$$
\lim_{u\to\infty} \max_{\tau_u} |F_u(x) - F(x)| = 0. \tag{2.5.5}
$$

We continue with the proof of Lemma 2.5.1 and postpone the proof of Lemma 2.5.2 to the end of this section. Recall that, for two random vectors $X, Y \in \mathbb{R}^m$, their covariance is defined as $\mathrm{Cov}(X,Y) := \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)^\top]$ and the variance matrix of X is defined as $\mathrm{Var}(X) := \mathrm{Cov}(X,X)$. The conditional mean of $(\xi_u(t), \eta_u(t))^\top$ given $X_1(s_u) = u -$

$\frac{x}{u}$, $X_2(t_u) = u - \frac{y}{u}$, is

$$\mathbb{E}\left(\begin{array}{c|c} \xi_u(t) & X_1(s_u) = u - \frac{x}{u} \\ \\ \eta_u(t) & X_2(t_u) = u - \frac{y}{u} \end{array}\right) = \mathbb{E}\begin{pmatrix} \xi_u(t) \\ \\ \eta_u(t) \end{pmatrix}$$

$$+ \operatorname{Cov}\left(\begin{pmatrix} \xi_u(t) \\ \\ \eta_u(t) \end{pmatrix}, \begin{pmatrix} X_1(s_u) \\ \\ X_2(t_u) \end{pmatrix}\right)\left(\operatorname{Var}\begin{pmatrix} X_1(s_u) \\ \\ X_2(t_u) \end{pmatrix}\right)^{-1}\begin{pmatrix} u - \frac{x}{u} \\ \\ u - \frac{y}{u} \end{pmatrix}$$

$$= \begin{pmatrix} -u^2 + x \\ \\ -u^2 + y \end{pmatrix} + \frac{u}{1 - r^2(|\tau_u|)}\begin{pmatrix} r_{11}(s_u + u^{-2/\alpha_1}t, s_u) & r(|\tau_u - u^{-2/\alpha_1}t|) \\ \\ r(|\tau_u + u^{-2/\alpha_2}t|) & r_{22}(t_u + u^{-2/\alpha_2}t, t_u) \end{pmatrix}$$

$$\times \begin{pmatrix} 1 & -r(|\tau_u|) \\ \\ -r(|\tau_u|) & 1 \end{pmatrix}\begin{pmatrix} u - \frac{x}{u} \\ \\ u - \frac{y}{u} \end{pmatrix}$$

$$\triangleq \begin{pmatrix} a_1(u) \\ \\ a_2(u) \end{pmatrix}, \tag{2.5.6}$$

where

$$\begin{aligned} a_1(u) = &- \frac{u^2(1 - r_{11}(s_u + u^{-2/\alpha_1}t, s_u)) - u^2(r(|\tau_u - u^{-2/\alpha_1}t|) - r(|\tau_u|))}{1 + r(|\tau_u|)} \\ &+ \frac{(x - yr(|\tau_u|))(1 - r_{11}(s_u + u^{-2/\alpha_1}t, s_u))}{1 - r^2(|\tau_u|)} \\ &+ \frac{(y - xr(|\tau_u|))(r(|\tau_u|) - r(|\tau_u - u^{-2/\alpha_1}t|))}{1 - r^2(|\tau_u|)} \end{aligned} \tag{2.5.7}$$

and

$$\begin{aligned} a_2(u) = &- \frac{u^2(1 - r_{22}(t_u + u^{-2/\alpha_2}t, t_u)) - u^2(r(|\tau_u + u^{-2/\alpha_2}t|) - r(|\tau_u|))}{1 + r(|\tau_u|)} \\ &+ \frac{(y - xr(|\tau_u|))(1 - r_{22}(t_u + u^{-2/\alpha_1}t, t_u))}{1 - r^2(|\tau_u|)} \end{aligned}$$

$$+\frac{(x-yr(|\tau_u|))(r(|\tau_u|)-r(|\tau_u+u^{-2/\alpha_2}t|))}{1-r^2(|\tau_u|)}. \tag{2.5.8}$$

Applying the mean value theorem twice, we see that for u large enough,

$$\begin{aligned}
&|r(|\tau_u+u^{-2/\alpha}t|)-r(|\tau_u|)| \le |u^{-2/\alpha}t|\cdot \max_{\substack{s \text{ is between} \\ |\tau_u| \text{ and } |\tau_u+u^{-2/\alpha}t|}} |r'(s)| \\
&\le |u^{-2/\alpha}t|\cdot \max_{|s|\le 2C\sqrt{\log u}/u} |r'(s)| \\
&\le |u^{-2/\alpha}t|\cdot \max_{|s|\le 2C\sqrt{\log u}/u} \left(|s|\cdot \max_{|t|\le|s|}|r''(t)|\right) \\
&\le 2C|t|\sqrt{\log u}\cdot u^{-1-2/\alpha}\cdot \max_{|t|\le 2C\sqrt{\log u}/u}|r''(t)| \\
&\le 4C|r''(0)||t|\sqrt{\log u}\cdot u^{-1-2/\alpha},
\end{aligned} \tag{2.5.9}$$

where the second inequality holds because of $u^{-2/\alpha}=o(\sqrt{\log u}/u)$, as $u\to\infty$ and the last inequality holds since $r''(\cdot)$ is continuous in a neighborhood of zero. Thus (2.5.9) implies that, as $u\to\infty$,

$$u^2|r(|\tau_u+u^{-2/\alpha}t|)-r(|\tau_u|)| \le 4C|r''(0)||t|\sqrt{\log u}\cdot u^{1-2/\alpha}\to 0, \tag{2.5.10}$$

where the convergence is uniform for all s_u and t_u that satisfy $|\tau_u|\le C\sqrt{\log u}/u$. We also notice that for $i=1,2$ and all $s\in\mathbb{R}^N$,

$$1-r_{ii}(s+u^{-2/\alpha}t,s)=c_iu^{-2}|t|^{\alpha_i}+o(u^{-2}), \text{ as } u\to\infty. \tag{2.5.11}$$

By (2.5.6), (2.5.10), and (2.5.11), we conclude that, as $u \to \infty$,

$$\mathbb{E}\left(\begin{array}{c|c} \xi_u(t) & X_1(s_u) = u - \frac{x}{u} \\ \\ \eta_u(t) & X_2(t_u) = u - \frac{y}{u} \end{array} \right) \to \begin{pmatrix} -\frac{c_1|t|^{\alpha_1}}{1+\rho} \\ \\ -\frac{c_2|t|^{\alpha_2}}{1+\rho} \end{pmatrix}, \tag{2.5.12}$$

where the convergence is uniform w.r.t. s_u and t_u satisfying $|\tau_u| \le C\sqrt{\log u}/u$.

Next, we consider the conditional covariance matrix of $(\xi_u(t) - \xi_u(s), \eta_u(t) - \eta_u(s))^\top$.

$$\begin{aligned} &\mathrm{Var}\left(\begin{pmatrix} \xi_u(t) - \xi_u(s) \\ \\ \eta_u(t) - \eta_u(s) \end{pmatrix} \middle| \begin{array}{c} X_1(s_u) \\ \\ X_2(t_u) \end{array} \right) \\ &= \mathrm{Var}\begin{pmatrix} \xi_u(t) - \xi_u(s) \\ \\ \eta_u(t) - \eta_u(s) \end{pmatrix} - \mathrm{Cov}\left(\begin{pmatrix} \xi_u(t) - \xi_u(s) \\ \\ \eta_u(t) - \eta_u(s) \end{pmatrix}, \begin{pmatrix} X_1(s_u) \\ \\ X_2(t_u) \end{pmatrix} \right) \\ &\quad \times \mathrm{Var}\begin{pmatrix} X_1(s_u) \\ \\ X_2(t_u) \end{pmatrix}^{-1} \mathrm{Cov}\left(\begin{pmatrix} \xi_u(t) - \xi_u(s) \\ \\ \eta_u(t) - \eta_u(s) \end{pmatrix}, \begin{pmatrix} X_1(s_u) \\ \\ X_2(t_u) \end{pmatrix} \right)^\top. \end{aligned} \tag{2.5.13}$$

Let $h_u(t,s) := r(|\tau_u + u^{-2/\alpha_2} t - u^{-2/\alpha_1} s|)$. Applying (2.5.10) and (2.5.11), we obtain

$$\begin{aligned} &\mathrm{Var}\begin{pmatrix} \xi_u(t) - \xi_u(s) \\ \\ \eta_u(t) - \eta_u(s) \end{pmatrix} \\ &= \begin{pmatrix} \begin{array}{l} 2u^2(1 - r_{11}(s_u + \\ \quad u^{-2/\alpha_1}s,\ s_u + u^{-2/\alpha_1}t)) \end{array} & \begin{array}{l} u^2(h_u(t,t) - h_u(s,t) \\ \quad -h_u(t,s) + h_u(s,s)) \end{array} \\ \\ \begin{array}{l} u^2(h_u(t,t) - h_u(s,t) \\ \quad -h_u(t,s) + h_u(s,s)) \end{array} & \begin{array}{l} 2u^2(1 - r_{22}(t_u + \\ \quad {}^{-2/\alpha_2}s,\ t_u + u^{-2/\alpha_2}t)) \end{array} \end{pmatrix} \end{aligned}$$

$$
= \begin{pmatrix} 2c_1|t-s|^{\alpha_1}(1+o(1)) & o(1) \\ o(1) & 2c_2|t-s|^{\alpha_2}(1+o(1)) \end{pmatrix}, \tag{2.5.14}
$$

where $o(1)$ converges to zero uniformly w.r.t. τ_u satisfying $|\tau_u| \le C\sqrt{\log u}/u$, as $u \to \infty$. Also, we have

$$
\begin{aligned}
&\operatorname{Cov}\left[\begin{pmatrix} \xi_u(t)-\xi_u(s) \\ \eta_u(t)-\eta_u(s) \end{pmatrix}, \begin{pmatrix} X_1(s_u) \\ X_2(t_u) \end{pmatrix}\right] \\
&= \begin{pmatrix} u(r_{11}(s_u+u^{-2/\alpha_1}t, s_u) & u(r(|\tau_u - u^{-2/\alpha_1}t|) \\ \quad -r_{11}(s_u+u^{-2/\alpha_1}s, s_u)) & \quad -r(|\tau_u - u^{-2/\alpha_1}s|)) \\ & \\ u(r(|\tau_u + u^{-2/\alpha_2}t|) & u(r_{22}(t_u+u^{-2/\alpha_2}t, t_u) \\ \quad -r(|\tau_u + u^{-2/\alpha_2}s|)) & \quad -r_{22}(t_u+u^{-2/\alpha_2}s, t_u)) \end{pmatrix} \\
&= \begin{pmatrix} o(1) & o(1) \\ o(1) & o(1) \end{pmatrix},
\end{aligned} \tag{2.5.15}
$$

as $u \to \infty$, and

$$
\operatorname{Var}\begin{pmatrix} X_1(s_u) \\ X_2(t_u) \end{pmatrix}^{-1} = \frac{1}{1-r^2(|\tau_u|)} \begin{pmatrix} 1 & -r(|\tau_u|) \\ -r(|\tau_u|) & 1 \end{pmatrix}. \tag{2.5.16}
$$

By (2.5.13) – (2.5.16), we conclude that as $u \to \infty$,

$$
\operatorname{Var}\left(\begin{pmatrix} \xi_u(t)-\xi_u(s) \\ \eta_u(t)-\eta_u(s) \end{pmatrix} \middle| \begin{matrix} X_1(s_u) \\ X_2(t_u) \end{matrix}\right) \to \begin{pmatrix} 2c_1|t-s|^{\alpha_1} & 0 \\ 0 & 2c_2|t-s|^{\alpha_2} \end{pmatrix}, \tag{2.5.17}
$$

where the convergence is uniform w.r.t. τ_u satisfying $|\tau_u| \le C\sqrt{\log u}/u$. Hence, the uniform convergence of f.d.d. in Lemma 2.5.1 follows from (2.5.12), (2.5.17) and Lemma 2.5.2.

Now we prove the second part of Lemma 2.5.1. The continuous mapping theorem (see, e.g., [Bil68], p. 30) can be used to prove (2.5.3) holds when s_u and t_u are fixed. Since we need to prove uniform convergence w.r.t. s_u and t_u, we use a discretization method instead. Let

$$f(u,x,y) := \mathbb{P}\Big(\max_{s\in\mathbb{S}} \xi_u(s) > x,\ \max_{t\in\mathbb{T}} \eta_u(t) > y \ \Big| \ X_1(s_u) = u - \frac{x}{u}, X_2(t_u) = u - \frac{y}{u}\Big) \tag{2.5.18}$$

and

$$f(x,y) := \mathbb{P}\Big(\max_{s\in\mathbb{S}} \xi(s) > x,\ \max_{t\in\mathbb{T}} \eta(t) > y\Big). \tag{2.5.19}$$

Without loss of generality, suppose that $\mathbb{S} = [a,b]^N$ and $\mathbb{T} = [c,d]^N$, where $a < b, c < d$. For any $\delta \in (0,1)$, let $m = \left\lfloor \frac{b-a}{\delta} \right\rfloor$, $n = \left\lfloor \frac{d-c}{\delta} \right\rfloor$ and let

$$\mathcal{S}_m := \big\{ s_{\mathbf{k}} \mid s_{\mathbf{k}} = (x_{k_1}, \dots, x_{k_N}),\ \mathbf{k} = (k_1, \dots, k_N) \in \{0,1,\dots,m+1\}^N \big\},$$
$$\mathcal{T}_n := \big\{ t_{\mathbf{l}} \mid t_{\mathbf{l}} = (y_{l_1}, \dots, y_{l_N}),\ \mathbf{l} = (l_1, \dots, l_N) \in \{0,1,\dots,n+1\}^N \big\},$$

where x_i, y_i are defined as

$$a = x_0 < x_1 < \cdots < x_m \le x_{m+1} = b,\ x_i = a + i\delta, i = 0,1,\dots,m,$$
$$c = y_0 < y_1 < \cdots < y_n \le y_{n+1} = d,\ \ y_i = c + i\delta, i = 0,1,\dots,n. \tag{2.5.20}$$

Then $[a,b]^N \times [c,d]^N$ can be divided into δ-cubes with vertices in $\mathcal{S}_m \times \mathcal{T}_n$.

The function $f(u,x,y)$ in (2.5.18) is bounded from below by

$$\begin{aligned} f_{m,n}(u,x,y) := \mathbb{P}\Big(\max_{s\in\mathcal{S}_m} \xi_u(s) > x,\ \max_{t\in\mathcal{T}_n} \eta_u(t) > y \ \Big| \\ X_1(s_u) = u - \frac{x}{u}, X_2(t_u) = u - \frac{y}{u} \Big) \end{aligned} \tag{2.5.21}$$

and is bounded from above by $g_{m,n}(u,x,y)$ which is defined as

$$\begin{aligned} &\mathbb{P}\Big(\max_{s\in\mathcal{S}_m} \xi_u(s) > x-\epsilon,\ \max_{t\in\mathcal{T}_n} \eta_u(t) > y-\epsilon \,\Big|\, X_1(s_u) = u - \frac{x}{u}, X_2(t_u) = u - \frac{y}{u} \Big) \\ &+ \mathbb{P}\Big(\max_{s\in\mathbb{S}} \xi_u(s) > x,\ \max_{s\in\mathcal{S}_m} \xi_u(s) \le x-\epsilon \,\Big|\, X_1(s_u) = u - \frac{x}{u}, X_2(t_u) = u - \frac{y}{u} \Big) \\ &+ \mathbb{P}\Big(\max_{t\in\mathbb{T}} \eta_u(t) > y,\ \max_{t\in\mathcal{T}_n} \eta_u(t) \le y-\epsilon \,\Big|\, X_1(s_u) = u - \frac{x}{u}, X_2(t_u) = u - \frac{y}{u} \Big) \\ &\triangleq f_{m,n}(u, x-\epsilon, y-\epsilon) + s_{m,n}(u,x,y) + t_{m,n}(u,x,y), \end{aligned} \tag{2.5.22}$$

where $\epsilon > 0$ is any small constant. Let

$$f_{m,n}(x,y) := \mathbb{P}\Big(\max_{s\in\mathcal{S}_m} \xi(s) > x,\ \max_{t\in\mathcal{T}_n} \eta(t) > y \Big). \tag{2.5.23}$$

Since the finite dimensional distributions of $(\xi_u(\cdot), \eta_u(\cdot))$ converge uniformly to those of $(\xi(\cdot), \eta(\cdot))$, we have

$$\lim_{u\to\infty} \max_{|\tau_u| \le C\sqrt{\log u}/u} |f_{m,n}(u,x,y) - f_{m,n}(x,y)| = 0. \tag{2.5.24}$$

The continuity of the trajectory of $(\xi(\cdot), \eta(\cdot))$ yields

$$\lim_{\substack{m\to\infty \\ n\to\infty}} f_{m,n}(x,y) = f(x,y). \tag{2.5.25}$$

By (2.5.24) and (2.5.25), we conclude

$$\lim_{\substack{m\to\infty \\ n\to\infty}} \lim_{u\to\infty} \max_{|\tau_u|\le C\sqrt{\log u}/u} |f_{m,n}(u,x,y) - f(x,y)| = 0. \tag{2.5.26}$$

Let us consider the conditional probability $s_{m,n}(u,x,y)$ in (2.5.22).

$$\begin{aligned}
&s_{m,n}(u,x,y) \\
&\le \mathbb{P}\bigg(\max_{|s-t|\le\delta} |\xi_u(s) - \xi_u(t)| > \epsilon \,\bigg|\, X_1(s_u) = u - \frac{x}{u},\ X_2(t_u) = u - \frac{y}{u} \bigg) \\
&\le \frac{1}{\epsilon}\mathbb{E}\left(\max_{|s-t|\le\delta} |\xi_u(s) - \xi_u(t)| \,\bigg|\, X_1(s_u) = u - \frac{x}{u},\ X_2(t_u) = u - \frac{y}{u} \right) \\
&= \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}_u}\bigg(\max_{|s-t|\le\delta} |x(s) - x(t)| \bigg),
\end{aligned} \tag{2.5.27}$$

where $\mathbb{P}_u$ is the probability measure on $(C(\mathbb{S}), \mathcal{B}(C(\mathbb{S}))$ defined as

$$\mathbb{P}_u(A) := \mathbb{P}\bigg(\xi_u(\cdot) \in A \,\bigg|\, X_1(s_u) = u - \frac{x}{u}, X_2(t_u) = u - \frac{y}{u} \bigg),$$

for all $A \in \mathcal{B}(C(\mathbb{S}))$ and $x(\cdot)$ is the coordinate random element on $(C(\mathbb{S}), \mathcal{B}(C(\mathbb{S})), \mathbb{P}_u)$, i.e., $x(t,\omega) = \omega(t),\ \forall \omega \in C(\mathbb{S})$ and $t \in \mathbb{S}$. Consider the canonical metric

$$\begin{aligned}
d_u(s,t) :&= \Big[\mathbb{E}_{\mathbb{P}_u}\big(|x(s) - x(t)|^2\big)\Big]^{1/2} \\
&= \left[\mathbb{E}\left(|\xi_u(s) - \xi_u(t)|^2 \,\bigg|\, X_1(s_u) = u - \frac{x}{u}, X_2(t_u) = u - \frac{y}{u} \right)\right]^{1/2}.
\end{aligned}$$

By (2.5.17), we have

$$d_u(s,t) \le 2\sqrt{c_1}|s-t|^{\alpha_1/2} \tag{2.5.28}$$

for u large enough and all s_u, t_u such that $|\tau_u| \le C\sqrt{\log u}/u$. If we choose $\gamma = \left(\frac{\epsilon}{2\sqrt{c_1}}\right)^{\frac{2}{\alpha_1}}$, then $d_u(s,t) < \epsilon$ for all $t, s \in \mathbb{S}$ with $|t-s| < \gamma$. Hence

$$N_{d_u}(\mathbb{S}, \epsilon) \le C\epsilon^{-2N/\alpha_1}, \tag{2.5.29}$$

where $N_{d_u}(\mathbb{S}, \epsilon)$ denotes the minimum number of d_u-balls with radius ϵ that are needed to cover $\mathbb{S}$. By Dudley's Theorem (see, e.g., Theorem 1.3.3 in [AT07]) and (2.5.28), we have

$$\mathbb{E}_{\mathbb{P}_u}\left(\max_{|s-t|\le\delta} |x(s)-x(t)|\right) \le K\int_0^{2\sqrt{c_1}\delta^{\alpha_1/2}} \sqrt{\log N_{d_u}(\mathbb{S},\epsilon)}\, d\epsilon, \tag{2.5.30}$$

where $K < \infty$ is a constant (which does not depend on δ) and, thanks to (2.5.29), the last integral goes to 0 as $\delta \to 0$ (or, equivalently, as $m \to \infty, n \to \infty$). By (2.5.27) and (2.5.30), we conclude that

$$\lim_{\substack{m\to\infty \\ n\to\infty}} \limsup_{u\to\infty} \max_{|\tau_u|\le C\sqrt{\log u}/u} |s_{m,n}(u,x,y)| = 0. \tag{2.5.31}$$

A similar argument shows that

$$\lim_{\substack{m\to\infty \\ n\to\infty}} \limsup_{u\to\infty} \max_{|\tau_u|\le C\sqrt{\log u}/u} |t_{m,n}(u,x,y)| = 0. \tag{2.5.32}$$

Since

$$|f(u,x,y) - f(x,y)| \le |f_{m,n}(u,x,y) - f(x,y)| + |g_{m,n}(u,x,y) - f(x,y)|$$

$$\le |f_{m,n}(u,x,y) - f(x,y)| + |f_{m,n}(u,x-\epsilon,y-\epsilon) - f(x-\epsilon,y-\epsilon)|$$

$$+\,|f(x-\epsilon,y-\epsilon)-f(x,y)|+|s_{m,n}(u,x,y)|+|t_{m,n}(u,x,y)|, \tag{2.5.33}$$

we combine (2.5.26), (2.5.31) and (2.5.32) to obtain

$$\begin{aligned}
&\limsup_{u\to\infty}\max_{|\tau_u|\le C\sqrt{\log u}/u}|f(u,x,y)-f(x,y)|\le|f(x-\epsilon,y-\epsilon)-f(x,y)|\\
&\quad+\lim_{\substack{m\to\infty\\ n\to\infty}}\limsup_{u\to\infty}\max_{|\tau_u|\le C\sqrt{\log u}/u}\Big(|f_{m,n}(u,x,y)-f(x,y)|+|s_{m,n}(u,x,y)|\\
&\quad+|t_{m,n}(u,x,y)|+|f_{m,n}(u,x-\epsilon,y-\epsilon)-f(x-\epsilon,y-\epsilon)|\Big)\\
&=|f(x-\epsilon,y-\epsilon)-f(x,y)|.
\end{aligned}$$

Since the last term $\to 0$ as $\epsilon\downarrow 0$, we have completed the proof of the second part of the lemma. $\square$

Now we are ready to prove the main lemmas in Section 3.

Proof of Lemma 2.4.1. Let $\phi(a,b)$ be the density of $(X_1(s_u),X_2(t_u))^\top$, i.e.,

$$\phi(a,b)=\frac{1}{2\pi\sqrt{1-r^2(|\tau_u|)}}\exp\left\{-\frac{1}{2}\frac{a^2-2r(|\tau_u|)ab+b^2}{1-r^2(|\tau_u|)}\right\}. \tag{2.5.34}$$

By conditioning and a change of variables, the LHS of (2.4.1) becomes

$$\begin{aligned}
&\mathbb{P}\Big(\max_{s\in s_u+u^{-2/\alpha_1}\mathbb{S}}X_1(s)>u,\max_{t\in t_u+u^{-2/\alpha_2}\mathbb{T}}X_2(t)>u\Big)\\
&=\int_{\mathbb{R}^2}\mathbb{P}\Big(\max_{s\in s_u+u^{-2/\alpha_1}\mathbb{S}}X_1(s)>u,\max_{t\in t_u+u^{-2/\alpha_2}\mathbb{T}}X_2(t)>u\;\Big|\;X_1(s_u)=u-\frac{x}{u},\\
&\qquad X_2(t_u)=u-\frac{y}{u}\Big)\phi\Big(u-\frac{x}{u},u-\frac{y}{u}\Big)u^{-2}dxdy\\
&=\frac{1}{2\pi\sqrt{1-r^2(|\tau_u|)}u^2}\exp\left(-\frac{u^2}{1+r(|\tau_u|)}\right)\int_{\mathbb{R}^2}f(u,x,y)\tilde{\phi}(u,x,y)dxdy,
\end{aligned} \tag{2.5.35}$$

where $f(u,x,y)$ is defined in (2.5.18) with $\xi_u(\cdot), \eta_u(\cdot)$ in (2.5.1), and where

$$\begin{aligned}&\tilde{\phi}(u,x,y)\\ &:= \exp\left\{-\frac{1}{2(1-r^2(|\tau_u|))}\Big(\frac{x^2+y^2}{u^2} - 2(1-r(|\tau_u|))(x+y) - 2r(|\tau_u|)\frac{xy}{u^2}\Big)\right\}.\end{aligned}$$

Since $\max_{|\tau_u|\le C\sqrt{\log u}/u} |r(|\tau_u|) - \rho| \to 0$ as $u \to \infty$, it is easy to check that

$$\max_{|\tau_u|\le C\sqrt{\log u}/u} \left|\tilde{\phi}(u,x,y) - e^{\frac{x+y}{1+\rho}}\right| \to 0, \quad \text{as} \quad u \to \infty. \tag{2.5.36}$$

Recall $H_\alpha(\cdot)$ in (2.2.1) and $f(x,y)$ in (2.5.19). Since $\xi(\cdot)$, $\eta(\cdot)$ are independent, and

$$\begin{aligned}\{\xi(t), t\in\mathbb{R}^N\} &\overset{d}{=} \left\{(1+\rho)\Big[\chi_1\Big(\Big(\frac{\sqrt{c_1}}{1+\rho}\Big)^{\frac{2}{\alpha_1}}t\Big) - \left|\Big(\frac{\sqrt{c_1}}{1+\rho}\Big)^{\frac{2}{\alpha_1}}t\right|^{\alpha_1}\Big], t\in\mathbb{R}^N\right\},\\ \{\eta(t)t\in\mathbb{R}^N\} &\overset{d}{=} \left\{(1+\rho)\Big[\chi_2\Big(\Big(\frac{\sqrt{c_2}}{1+\rho}\Big)^{\frac{2}{\alpha_2}}t\Big) - \left|\Big(\frac{\sqrt{c_2}}{1+\rho}\Big)^{\frac{2}{\alpha_2}}t\right|^{\alpha_2}\Big], t\in\mathbb{R}^N\right\},\end{aligned}$$

where $\overset{d}{=}$ means equality of all finite dimensional distributions, we have

$$\begin{aligned}&\int_{\mathbb{R}^2} f(x,y)e^{\frac{x+y}{1+\rho}}dxdy\\ &= \int_{\mathbb{R}} e^{\frac{x}{1+\rho}}\mathbb{P}\Big(\max_{s\in\mathbb{S}}\xi(s) > x\Big)dx \int_{\mathbb{R}} e^{\frac{y}{1+\rho}}\mathbb{P}\Big(\max_{t\in\mathbb{T}}\eta(t) > y\Big)dy\\ &= (1+\rho)^2 H_{\alpha_1}\Bigg(\frac{c_1^{1/\alpha_1}\mathbb{S}}{(1+\rho)^{\frac{2}{\alpha_1}}}\Bigg)H_{\alpha_2}\Bigg(\frac{c_2^{1/\alpha_2}\mathbb{T}}{(1+\rho)^{\frac{2}{\alpha_2}}}\Bigg).\end{aligned} \tag{2.5.37}$$

By (2.5.35) and (2.5.37), to conclude the lemma, it suffices to prove

$$\lim_{u\to\infty}\int_{\mathbb{R}^2}\max_{|\tau_u|\le C\sqrt{\log u}/u}\left|f(u,x,y)\tilde{\phi}(u,x,y) - f(x,y)e^{\frac{x+y}{1+\rho}}\right| dxdy = 0. \tag{2.5.38}$$

Firstly, applying Lemma 2.5.1 together with (2.5.36), we have

$$\max_{|\tau u| \le C\sqrt{\log u}/u} \left| f(u,x,y)\tilde{\phi}(u,x,y) - f(x,y)e^{\frac{x+y}{1+\rho}} \right| \to 0, \text{ as } u \to \infty. \tag{2.5.39}$$

Secondly, as in [LP00], we can find an integrable dominating function $g \in L(\mathbb{R}^2)$ such that for u large enough,

$$\max_{|\tau u| \le C\sqrt{\log u}/u} \left| f(u,x,y)\tilde{\phi}(u,x,y) - f(x,y)e^{\frac{x+y}{1+\rho}} \right| \le g(x,y). \tag{2.5.40}$$

Therefore, (2.5.38) follows from the dominated convergence theorem. This finishes the proof.

□

Proof of Lemma 2.4.2. We first claim that for any compact sets $\mathbb{S}$ and $\mathbb{T}$, the identity

$$H_\alpha(\mathbb{S}) + H_\alpha(\mathbb{T}) - H_\alpha(\mathbb{S} \cup \mathbb{T}) = H_\alpha(\mathbb{S}, \mathbb{T}) \tag{2.5.41}$$

holds. Indeed, if we let $X = \sup_{t\in\mathbb{S}}(\chi(t) - |t|^\alpha)$ and $Y = \sup_{t\in\mathbb{T}}(\chi(t) - |t|^\alpha)$, then

$$\begin{aligned} H_\alpha(\mathbb{S}) + H_\alpha(\mathbb{T}) - H_\alpha(\mathbb{S} \cup \mathbb{T}) &= \mathbb{E}\big(e^X\big) + \mathbb{E}\big(e^Y\big) - \mathbb{E}\big(e^{\max(X,Y)}\big) \\ = \mathbb{E}\big(e^X 1_{\{X<Y\}}\big) + \mathbb{E}\big(e^Y 1_{\{X\ge Y\}}\big) &= \mathbb{E}\big(e^{\min(X,Y)}\big) = H_\alpha(\mathbb{S}, \mathbb{T}). \end{aligned}$$

Now let $\mathbb{T}_1 = [0,T]^N$, $\mathbb{T}_2 = [\mathbf{m}T, (\mathbf{m}+1)T]$ and $\mathbb{T}_3 = [\mathbf{n}T, (\mathbf{n}+1)T]$. Consider the events

$$A = \left\{ \max_{s \in s_u + u^{-2/\alpha_1}\mathbb{T}_1} X_1(s) > u \right\}, \quad B = \left\{ \max_{s \in s_u + u^{-2/\alpha_1}\mathbb{T}_2} X_1(s) > u \right\},$$
$$C = \left\{ \max_{t \in t_u + u^{-2/\alpha_2}\mathbb{T}_1} X_2(t) > u \right\}, \quad D = \left\{ \max_{t \in t_u + u^{-2/\alpha_2}\mathbb{T}_3} X_2(t) > u \right\}.$$

It is easy to check that the LHS of (2.4.2) is equal to

$$
\begin{aligned}
&\mathbb{P}(A \cap B \cap C \cap D) \\
&= [\mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}((A \cup B) \cap C)] \\
&+ [\mathbb{P}(A \cap D) + \mathbb{P}(B \cap D) - \mathbb{P}((A \cup B) \cap D)] \\
&- [\mathbb{P}(A \cap (C \cup D)) + \mathbb{P}(B \cap (C \cup D)) - \mathbb{P}((A \cup B) \cap (C \cup D))]. \qquad (2.5.42)
\end{aligned}
$$

Let $R(u) = \frac{(1+\rho)^2}{2\pi\sqrt{1-\rho^2}} u^{-2} \exp\left(-\frac{u^2}{1+r(|\tau_u|)}\right)$ and $q_{\alpha,c} = \frac{(1+\rho)^{2/\alpha}}{c^{1/\alpha}}$. By Lemma 2.4.1, we have

$$
\begin{aligned}
\mathbb{P}(A \cap C) &= R(u) H_{\alpha_1}\left(\frac{\mathbb{T}_1}{q_{\alpha_1,c_1}}\right) H_{\alpha_2}\left(\frac{\mathbb{T}_1}{q_{\alpha_2,c_2}}\right)(1+\gamma_1(u)), \\
\mathbb{P}(B \cap C) &= R(u) H_{\alpha_1}\left(\frac{\mathbb{T}_2}{q_{\alpha_1,c_1}}\right) H_{\alpha_2}\left(\frac{\mathbb{T}_1}{q_{\alpha_2,c_2}}\right)(1+\gamma_2(u)), \\
\mathbb{P}((A \cup B) \cap C) &= R(u) H_{\alpha_1}\left(\frac{\mathbb{T}_1 \cup \mathbb{T}_2}{q_{\alpha_1,c_1}}\right) H_{\alpha_2}\left(\frac{\mathbb{T}_1}{q_{\alpha_2,c_2}}\right)(1+\gamma_3(u)),
\end{aligned}
$$

where, for $i = 1, 2, 3$, $\gamma_i(u) \to 0$ uniformly w.r.t. τ_u satisfying $|\tau_u| \le C\sqrt{\log u}/u$, as $u \to \infty$. These, together with (2.5.41), imply

$$
\begin{aligned}
&\mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}((A \cup B) \cap C) \\
&= R(u) H_{\alpha_2}\left(\frac{\mathbb{T}_1}{q_{\alpha_2,c_2}}\right) H_{\alpha_1}\left(\frac{\mathbb{T}_1}{q_{\alpha_1,c_1}}, \frac{\mathbb{T}_2}{q_{\alpha_1,c_1}}\right)(1+o(1)). \qquad (2.5.43)
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&\mathbb{P}(A \cap D) + \mathbb{P}(B \cap D) - \mathbb{P}((A \cup B) \cap D) \\
&= R(u) H_{\alpha_2}\left(\frac{\mathbb{T}_3}{q_{\alpha_2,c_2}}\right) H_{\alpha_1}\left(\frac{\mathbb{T}_1}{q_{\alpha_1,c_1}}, \frac{\mathbb{T}_2}{q_{\alpha_1,c_1}}\right)(1+o(1)) \qquad (2.5.44)
\end{aligned}
$$

and

$$\begin{aligned}
&\mathbb{P}(A \cap (C \cup D)) + \mathbb{P}(B \cap (C \cup D)) - \mathbb{P}((A \cup B) \cap (C \cup D)) \\
&= R(u) H_{\alpha_2}\left(\frac{\mathbb{T}_1 \cup \mathbb{T}_3}{q_{\alpha_2,c_2}}\right) H_{\alpha_1}\left(\frac{\mathbb{T}_1}{q_{\alpha_1,c_1}}, \frac{\mathbb{T}_2}{q_{\alpha_1,c_1}}\right)(1+o(1)). \qquad (2.5.45)
\end{aligned}$$

By (2.5.42) – (2.5.45), we have

$$\begin{aligned}
&\mathbb{P}(A \cap B \cap C \cap D) \\
&= R(u) H_{\alpha_1}\left(\frac{\mathbb{T}_1}{q_{\alpha_1,c_1}}, \frac{\mathbb{T}_2}{q_{\alpha_1,c_1}}\right) H_{\alpha_2,c_2}\left(\frac{\mathbb{T}_1}{q_{\alpha_2,c_2}}, \frac{\mathbb{T}_3}{q_{\alpha_2,c_2}}\right)(1+o(1)),
\end{aligned}$$

which conclude the lemma. □

Proof of Lemma 2.4.3. Let $f(|t|) = \frac{1}{1+r(|t|)}$. Recall $\tau_{\mathbf{kl}}$ defined in (2.4.9) and $|\tau_{\mathbf{kl}}| \leq 2\delta(u)$, when u is large. By Taylor's expansion,

$$f(|\tau_{\mathbf{kl}}|) = f(0) + \frac{1}{2} f''(0) |\tau_{\mathbf{kl}}|^2 (1 + \gamma_{\mathbf{kl}}(u)),$$

where $f(0) = \frac{1}{1+\rho}$, $f''(0) = \frac{-r''(0)}{(1+\rho)^2}$ and, as $u \to \infty$, $\gamma_{\mathbf{kl}}(u)$ converges to zero uniformly w.r.t. all $(\mathbf{k}, \mathbf{l}) \in \mathcal{C}$. Therefore, for any $\epsilon > 0$, we have

$$\sum_{(\mathbf{k},\mathbf{l}) \in \mathcal{C}} e^{-\frac{1}{2} f''(0)(1+\epsilon) u^2 |\tau_{\mathbf{kl}}|^2} \leq h(u) \leq \sum_{(\mathbf{k},\mathbf{l}) \in \mathcal{C}} e^{-\frac{1}{2} f''(0)(1-\epsilon) u^2 |\tau_{\mathbf{kl}}|^2} \qquad (2.5.46)$$

when u is large enough. For $a > 0$, let

$$h(u, a) := \sum_{(\mathbf{k},\mathbf{l}) \in \mathcal{C}} e^{-a u^2 |\tau_{\mathbf{kl}}|^2}. \qquad (2.5.47)$$

In order to prove (2.4.12), it suffices to prove that

$$\lim_{u\to\infty} u^N d_1^N(u) d_2^N(u) h(u,a) = \left(\frac{\pi}{a}\right)^{\frac{N}{2}} mes_N(A_1 \cap A_2). \tag{2.5.48}$$

To this end, we write

$$\begin{aligned} &u^N d_1^N(u) d_2^N(u) h(u,a) \\ &= \frac{1}{u^N} \sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}} e^{-a\sum_{j=1}^N (l_j u d_2(u) - k_j u d_1(u))^2} \cdot (ud_1(u))^N (ud_2(u))^N. \end{aligned} \tag{2.5.49}$$

Let

$$\begin{aligned} p(u) &:= \frac{1}{u^N} \sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}} \min_{(s,t)\in u\Delta_{\mathbf{k}}^{(1)} \times u\Delta_{\mathbf{l}}^{(2)}} e^{-a|t-s|^2} \cdot (ud_1(u))^N (ud_2(u))^N, \\ q(u) &:= \frac{1}{u^N} \sum_{(\mathbf{k},\mathbf{l})\in\mathcal{C}} \max_{(s,t)\in u\Delta_{\mathbf{k}}^{(1)} \times u\Delta_{\mathbf{l}}^{(2)}} e^{-a|t-s|^2} \cdot (ud_1(u))^N (ud_2(u)^N. \end{aligned}$$

It follows from (2.5.49) that

$$p(u) \le u^N d_1^N(u) d_2^N(u) h(u,a) \le q(u), \tag{2.5.50}$$

and

$$p(u) \le \frac{1}{u^N} \int_{\substack{s\in uA_1, t\in uA_2 \\ |t-s|\le C\sqrt{\log u}}} e^{-a|t-s|^2} dtds \le q(u). \tag{2.5.51}$$

Observe that

$$\frac{1}{u^N} \iint_{\substack{s\in uA_1, t\in uA_2 \\ |t-s|\le C\sqrt{\log u}}} e^{-a|t-s|^2} dtds = \frac{1}{u^N} \iint_{\substack{y\in uA_1, x+y\in uA_2 \\ |x|\le C\sqrt{\log u}}} e^{-a|x|^2} dxdy$$

$$= \frac{1}{u^N} \int_{|x| \le C\sqrt{\log u}} e^{-a|x|^2} dx \int_{\mathbb{R}^N} 1_{\{y \in uA_1 \cap (uA_2 - x)\}} dy$$
$$= \int_{|x| \le C\sqrt{\log u}} e^{-a|x|^2} dx \int_{\mathbb{R}^N} 1_{\{z \in A_1 \cap (A_2 - x/u)\}} dz$$
$$\to \; mes_N(A_1 \cap A_2) \int_{\mathbb{R}^N} e^{-a|x|^2} dx = \Big(\frac{\pi}{a}\Big)^{\frac{N}{2}} mes_N(A_1 \cap A_2), \tag{2.5.52}$$

as $u \to \infty$, where the convergence holds by the dominated convergence theorem. Indeed, $\int_{\mathbb{R}^N} 1_{\{z \in A_1 \cap (A_2 - x/u)\}} dz$ is bounded by $\max_{|\epsilon|<1} mes_N(A_1 \cap (A_2 - \epsilon))$ uniformly for $|x| \le C\sqrt{\log u}$ when u is large enough.

It follows from (2.5.50)–(2.5.50) that, for concluding (2.5.48), it remains to verify

$$D(u) := q(u) - p(u) \to 0, \text{ as } u \to \infty. \tag{2.5.53}$$

Define

$$\hat{\mathcal{D}} := \Big\{(s,t) \in A_1 \times A_2 : \, |t-s| \le \delta(u) + \sqrt{N} d_1(u) + \sqrt{N} d_2(u)\Big\}. \tag{2.5.54}$$

By the definition of $\mathcal{C}$ in (2.4.7), we see that $\mathcal{D} \subseteq \bigcup_{(\mathbf{k},\mathbf{l}) \in \mathcal{C}} \Delta_{\mathbf{k}}^{(1)} \times \Delta_{\mathbf{l}}^{(2)} \subseteq \hat{\mathcal{D}}$. Since $d_1(u) = o(\delta(u))$ and $d_2(u) = o(\delta(u))$ as $u \to \infty$, the set $\hat{\mathcal{D}}$ is a subset of $\tilde{\mathcal{D}} := \{(s,t) \in A_1 \times A_2 : |t-s| \le 2\delta(u)\}$ when u is large.

Write $D(u)$ in (2.5.53) as a sum over $(\mathbf{k},\mathbf{l}) \in \mathcal{C}$. To estimate the cardinality of $\mathcal{C}$, we notice that

$$mes_{2N}(\tilde{\mathcal{D}}) = \iint_{s \in A_1, t \in A_2} 1_{\{|t-s| \le 2\delta(u)\}} ds dt \tag{2.5.55}$$
$$= \int_{|x| \le 2\delta(u)} \int_{y \in A_1 \cap (A_2 - x)} dy dx \le K\, \delta(u)^N, \tag{2.5.56}$$

for all u large enough, where $K = 2^{N+1}\pi^{N/2}\Gamma^{-1}(N/2)\max_{|\epsilon|\le 1} mes_N(A_1 \cap (A_2 - \epsilon))$. Hence, for large u, the number of summands in (2.5.49) is bounded by

$$\#\{(\mathbf{k},\mathbf{l}) \mid (\mathbf{k},\mathbf{l}) \in \mathcal{C}\} \le \frac{mes_{2N}(\tilde{\mathcal{D}})}{mes_{2N}(\Delta_{\mathbf{k}}^{(1)} \times \Delta_{\mathbf{l}}^{(2)})} \le \frac{K\,\delta(u)^N}{d_1^N(u) d_2^N(u)}. \tag{2.5.57}$$

Next, by applying the inequality $e^{-x} - e^{-y} \le y - x$ for $y \ge x > 0$ to each summand in $D(u)$, we obtain

$$\begin{aligned}
&\max_{(s,t)\in u\Delta_{\mathbf{k}}^{(1)}\times u\Delta_{\mathbf{l}}^{(2)}} e^{-a|t-s|^2} - \min_{(s,t)\in u\Delta_{\mathbf{k}}^{(1)}\times u\Delta_{\mathbf{l}}^{(2)}} e^{-a|t-s|^2} \\
&\le a\left(\max_{(s,t)\in u\Delta_{\mathbf{k}}^{(1)}\times u\Delta_{\mathbf{l}}^{(2)}} |t-s|^2 - \min_{(s,t)\in u\Delta_{\mathbf{k}}^{(1)}\times u\Delta_{\mathbf{l}}^{(2)}} |t-s|^2\right) \\
&= a\max\big(|t-s| + |t_1 - s_1|\big)\big(|t-s| - |t_1 - s_1|\big),
\end{aligned} \tag{2.5.58}$$

where the last maximum is taken over $(s,t,s_1,t_1) \in u\Delta_{\mathbf{k}}^{(1)} \times u\Delta_{\mathbf{l}}^{(2)} \times u\Delta_{\mathbf{k}}^{(1)} \times u\Delta_{\mathbf{l}}^{(2)}$

Since $|t-s| \le 2\delta(u)$ for all $(t,s) \in u\Delta_{\mathbf{k}}^{(1)} \times u\Delta_{\mathbf{l}}^{(2)}$ when u is large, the inequality $\big||t-s| - |t_1 - s_1|\big| \le |t - t_1| + |s - s_1|$ implies that (2.5.58) is at most

$$4a\sqrt{N}u^2\delta(u)\big(d_1(u) + d_2(u)\big) \tag{2.5.59}$$

when u is large enough. By (2.5.59) and (2.5.57), we can verify that

$$\begin{aligned}
D(u) &\le \frac{1}{u^N}\frac{K(\delta(u))^N}{d_1^N(u) d_2^N(u)} 4a\sqrt{N}u^2\delta(u)\big(d_1(u)+d_2(u)\big)\big(ud_1(u)\big)^N\big(ud_2(u)\big)^N \\
&\le C_0(\log u)^{\frac{N+1}{2}}\big(u^{1-\frac{2}{\alpha_1}} + u^{1-\frac{2}{\alpha_2}}\big) \to 0, \quad \text{as} \quad u \to \infty.
\end{aligned}$$

Therefore (2.5.48) holds. Similarly, we can check that the same statement holds while changing the set $\mathcal{C}$ to $\mathcal{C}^\circ$. $\square$

Proof of Lemma 2.4.4. Inequality (2.4.22) holds immediately by Lemma 6.2 in [Pit96]. Hence we only consider the case when $\mathbf{m} \neq \mathbf{0}$. Suppose that $\{X(t), t \in \mathbb{R}^N\}$ is a real valued continuous Gaussian process with $\mathbb{E}[X(t)] = 0$ and covariance function $r(t)$ satisfying $r(t) = 1 - |t|^\alpha + o(|t|^\alpha)$ for a constant $\alpha \in (0, 2)$. Applying Lemma 6.1 in [Pit96], we see that for any $S > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{t\in u^{-2/\alpha}[0,S]^N} X(t) > u, \max_{t\in u^{-2/\alpha}[\mathbf{m}S,(\mathbf{m}+1)S]} X(t) > u\Big) \\
&= \mathbb{P}\Big(\max_{t\in u^{-2/\alpha}[0,S]^N} X(t) > u\Big) + \mathbb{P}\Big(\max_{t\in u^{-2/\alpha}[\mathbf{m}S,(\mathbf{m}+1)S]} X(t) > u\Big) \\
&\qquad - \mathbb{P}\Big(\max_{t\in u^{-2/\alpha}([0,S]^N\cup[\mathbf{m}S,(\mathbf{m}+1)S])} X(t) > u)\Big) \\
&= \Big(H_\alpha([0,S]^N) + H_\alpha([\mathbf{m}S,(\mathbf{m}+1)S]) - H_\alpha([0,S]^N \cup [\mathbf{m}S,(\mathbf{m}+1)S])\Big) \\
&\qquad \times \frac{1}{\sqrt{2\pi}u} e^{-\frac{1}{2}u^2}(1+o(1)) \\
&= H_\alpha([0,S]^N, [\mathbf{m}S,(\mathbf{m}+1)S]) \frac{1}{\sqrt{2\pi}u} e^{-\frac{1}{2}u^2}(1+o(1)), \quad \text{as} \quad u \to \infty, \qquad (2.5.60)
\end{aligned}$$

where the last equality holds thanks to (2.5.41).

On the other hand, by applying Lemma 6.3 in [Pit96] and the inequality $\inf_{s\in[0,1]^N,\, t\in[\mathbf{m},\mathbf{m}+1]} |s-t| \geq |m_{i_0}| - 1$ (recall that i_0 is defined in Lemma 2.4.4), we have

$$\begin{aligned}
&\mathbb{P}\Big(\max_{t\in u^{-2/\alpha}[0,S]^N} X(t) > u, \max_{t\in u^{-2/\alpha}[\mathbf{m}S,(\mathbf{m}+1)S]} X(t) > u\Big) \\
&\leq C_0 S^{2N} \frac{1}{\sqrt{2\pi}u} e^{-\frac{1}{2}u^2} \exp\Big(-\frac{1}{8}(|m_{i_0}| - 1)^\alpha S^\alpha\Big) \qquad (2.5.61)
\end{aligned}$$

for all u large enough. It follows from (2.5.60) and (2.5.61) that

$$H_\alpha([0,S]^N, [\mathbf{m}S, (\mathbf{m}+1)S]) \le C_0 S^{2N} \exp\left(-\frac{1}{8}(|m_{i_0}|-1)^\alpha S^\alpha\right), \tag{2.5.62}$$

which implies (2.4.24) by letting $S = \frac{c^{1/\alpha}T}{(1+\rho)^{2/\alpha}}$.

When $|m_{i_0}| = 1$, the above upper bound is not sharp. Instead, we derive (2.4.23) in Lemma 2.4.4 as follows. For concreteness, suppose that $i_0 = N$ and $m_N = 1$. By applying Lemmas 6.1 - 6.3 in [Pit96], we have

$$\begin{aligned}
&\mathbb{P}\Big(\max_{t\in u^{-2/\alpha}[0,S]^N} X(t) > u, \max_{t\in u^{-2/\alpha}[\mathbf{m}S,(\mathbf{m}+1)S]} X(t) > u\Big)\\
&\le \mathbb{P}\Big(\max_{t\in u^{-2/\alpha}(\prod_{j=1}^{N-1}[m_jS,(m_j+1)S]\times[S,S+\sqrt{S}])} X(t) > u\Big)\\
&+ \mathbb{P}\Big(\max_{t\in u^{-2/\alpha}[0,S]^N} X(t) > u, \max_{t\in u^{-2/\alpha}(\prod_{j=1}^{N-1}[m_jS,(m_j+1)S]\times[S+\sqrt{S},2S+\sqrt{S}])} X(t) > u\Big)\\
&\le C_0 S^{N-\frac{1}{2}} \frac{1}{\sqrt{2\pi}u} e^{-\frac{1}{2}u^2} + C_0 S^{2N} \frac{1}{\sqrt{2\pi}u} e^{-\frac{1}{2}u^2} e^{-\frac{1}{8}S^{\alpha/2}}\\
&\le C_0 S^{N-\frac{1}{2}} \frac{1}{\sqrt{2\pi}u} e^{-\frac{1}{2}u^2}
\end{aligned} \tag{2.5.63}$$

for u and S large. Hence, when $|m_{i_0}| = 1$, we have

$$H_\alpha([0,S]^N, [\mathbf{m}S, (\mathbf{m}+1)S]) \le C_0 S^{N-\frac{1}{2}} \tag{2.5.64}$$

for large S. This implies (2.4.23) by letting $S = \frac{c^{1/\alpha}T}{(1+\rho)^{2/\alpha}}$.

Notice that

$$\#\{m \in \mathbb{Z}^N \mid \max_{1\le i\le N} |m_i| = k\} = (2k+1)^N - (2k-1)^N, \ k = 1, 2, ... \tag{2.5.65}$$

By (2.4.23), (2.4.24) and the fact $\int_T^\infty x^N e^{-ax}dx \sim \frac{1}{a}T^N e^{-aT}$ as $T \to \infty$, we have

$$
\begin{aligned}
&\sum_{\mathbf{m}\neq \mathbf{0}} \mathcal{H}_{\alpha,c}(\mathbf{m}) = \sum_{k=1}^{\infty} \sum_{|m_{i_0}|=k} \mathcal{H}_{\alpha,c}(\mathbf{m}) \\
&\leq C_0(3^N-1)T^{N-\frac{1}{2}} + C_0 \sum_{k=2}^{\infty} [(2k+1)^N - (2k-1)^N] T^{2N} e^{-\frac{c}{8(1+\rho)^2}(k-1)^\alpha T^\alpha} \\
&\leq C_0(3^N-1)T^{N-\frac{1}{2}} + C_0 T^{2N} \int_1^\infty x^N e^{-\frac{c}{8(1+\rho)^2}x^\alpha T^\alpha} dx \leq C_0 T^{N-\frac{1}{2}}
\end{aligned}
$$

for T large enough. This completes the proof of Lemma 2.4.4. $\square$

Proof of Lemma 2.4.5. The proof is similar to that of Lemma 2.4.3. Indeed, we only need to modify (2.5.49) and (2.5.55) in the proof of Lemma 2.4.3. For any $y = (y_1, ..., y_N) \in \mathbb{R}^N$ and $1 \leq i \leq j \leq N$, let $y_{i:j} = (y_i, ..., y_j)$. On one hand, with a different scaling, $h(u,a)$ in (2.5.49) has the following asymptotics:

$$
\begin{aligned}
&u^{2N-M} d_1^N(u) d_2^N(u) h(u,a) \approx \frac{1}{u^M} \iint_{\substack{y\in uA_1, x+y\in uA_2 \\ |x|\leq C\sqrt{\log u}}} e^{-a|x|^2} dx dy \\
&= \frac{1}{u^M} \int_{|x|\leq C\sqrt{\log u}} e^{-a|x|^2} \Bigg(\int_{\mathbb{R}^M} 1_{\{y_{1:M}\in uA_{1,M}\cap(uA_{2,M}-x_{1:M})\}} dy_{1:M} \\
&\quad \times \prod_{j=M+1}^{N} \int_{\mathbb{R}} 1_{\{y_j\in[uS_j,uT_j]\cap[uT_j-x_j,uR_j-x_j]\}} dy_j \Bigg) dx \\
&= \int_{|x|\leq C\sqrt{\log u}} e^{-a|x|^2} \prod_{j=M+1}^{N} x_j 1_{\{x_j>0\}} \Bigg(\int_{\mathbb{R}^M} 1_{\{z_{1:M}\in A_{1,M}\cap(A_{2,M}-x_{1:M}/u)\}} dz_{1:M} \Bigg) dx \\
&\to \; mes_M(A_{1,M}\cap A_{2,M}) \int_{\mathbb{R}^M} e^{-a|x_{1:M}|^2} dx_{1:M} \prod_{j=M+1}^{N} \int_0^\infty x_j e^{-ax_j^2} dx_j \\
&= 2^{M-N}\pi^{M/2} a^{M/2-N} mes_M(A_{1,M}\cap A_{2,M}), \qquad (2.5.66)
\end{aligned}
$$

as $u \to \infty$. On the other hand, when u is large enough, $mes_{2N}(\tilde{\mathcal{D}})$ defined in (2.5.55) can be bounded from above by

$$
\begin{aligned}
mes_{2N}(\tilde{\mathcal{D}}) &= \iint_{s\in A_1, t\in A_2} 1_{\{|t-s|\le 2\delta(u)\}} ds dt \\
&= \int_{|x|\le 2\delta(u)} \Big(\int_{y_{1:M}\in A_{1,M}\cap(A_{2,M}-x_{1:M})} dy_{1:M} \Big) \prod_{j=M+1}^{N} x_j 1_{\{x_j>0\}} dx \\
&= \delta(u)^{2N-M} \int_{|z|\le 2} \Big(\int_{y_{1:M}\in A_{1,M}\cap(A_{2,M}-z_{1:M}\delta(u))} dy_{1:M} \Big) \prod_{j=M+1}^{N} z_j 1_{\{z_j>0\}} dz \\
&\le K\, \delta(u)^{2N-M}, \qquad (2.5.67)
\end{aligned}
$$

where $K = \max_{|\epsilon|\le 1} mes_M(A_{1,M}\cap(A_{2,M}-\epsilon)) \int_{|z|\le 2} \prod_{j=M+1}^{N} z_j 1_{\{z_j>0\}} dz$.

By (2.5.66) and (2.5.67), (2.4.36) can be obtained through the same argument in the proof of Lemma 2.4.3. We omit the details. □

We end this section with the proof of Lemma 2.5.2.

Proof of Lemma 2.5.2. Let $f_{u,\tau_u}(\cdot)$ and $f(\cdot)$ be the density function of $X(u,\tau_u)$ and X, respectively. It suffices to prove that for all $x \in \mathbb{R}^N$,

$$
\int_{\{y\le x\}} f(y) \max_{\tau_u} \left| \frac{f_{u,\tau_u(y)}}{f(y)} - 1 \right| dy \to 0, \text{ as } u \to \infty, \qquad (2.5.68)
$$

where $\{y \le x\} = \prod_{i=1}^{N}(-\infty, x_i]$.

First, we will find an upper bound for $\max_{\tau_u} |f_{u,\tau_u}(y)/f(y) - 1|$. For any $\epsilon > 0$, define

$$
\begin{aligned}
\Gamma(u,\tau_u) &= (\gamma_{ij}(u,\tau_u))_{i,j=1,\ldots n} := \frac{1}{\epsilon}(\Sigma(u,\tau_u) - \Sigma) \\
e(u,\tau_u) &= (e_i(u,\tau_u))_{i=1,\ldots n} := \frac{1}{\epsilon}(\mu(u,\tau_u) - \mu).
\end{aligned}
$$

By Assumption (2.5.4), there exists a constant $U > 0$ such that for all $u > U$,

$$\max_{\tau_u} |\mu_j(u,\tau_u) - \mu_j| < \epsilon, \quad \max_{\tau_u} |\sigma_{ij}(u,\tau_u) - \sigma_{ij}| < \epsilon, \quad i,j = 1,\dots,n,$$

which implies $|\gamma_{ij}(u,\tau_u)| \leq 1$ and $|e_i(u,\tau_u)| \leq 1$ for $u > U$.

Let $\Sigma^{-1} = (v_{ij})_{i,j=1,\dots,n}$ be the inverse of Σ. When ϵ is small, the determinant of $\Sigma(u,\tau_u)$ satisfies

$$|\Sigma(u,\tau_u)| = |\Sigma + \epsilon\Gamma(u,\tau_u)| = |\Sigma|(1 + \epsilon \mathrm{tr}(\Sigma^{-1}\Gamma(u,\tau_u)) + O(\epsilon^2)),$$

where $O(\epsilon^2)/\epsilon^2$ is uniformly bounded w.r.t. τ_u for large u (see, e.g., [MN07], p. 169). Hence we have

$$\left| \frac{|\Sigma(u,\tau_u)|}{|\Sigma|} - 1 \right| \leq 2\epsilon |\mathrm{tr}(\Sigma^{-1}\Gamma(u,\tau_u))| \leq 2\epsilon \sum_{i,j} |v_{ij}|. \tag{2.5.69}$$

Since $|\gamma_{ij}(u,\tau_u)| \leq 1,\ \forall i,j = 1,...,n$ for large u, as $\epsilon \to 0$, the inverse of $\Sigma(u,\tau_u)$ can be written as

$$\Sigma(u,\tau_u)^{-1} = \Sigma^{-1} - \epsilon\Sigma^{-1}\Gamma(u,\tau_u)\Sigma^{-1} + O(\epsilon^2),$$

where $O(\epsilon^2)/\epsilon^2$ is a matrix whose entries are uniformly bounded and independent of τ_u for large u (see, e.g., [Mey00], p. 618). Hence,

$$\begin{aligned} d_{u,\tau_u}(y) :=& -\frac{1}{2}\Big[(y - \mu(u,\tau_u))^\top \Sigma^{-1}(u,\tau_u)(y - \mu(u,\tau_u)) - (y-\mu)^\top \Sigma^{-1}(y-\mu)\Big] \\ =& -\frac{1}{2}(y-\mu)^\top \big(-\epsilon\Sigma^{-1}\Gamma(u,\tau_u)\Sigma^{-1} + O(\epsilon^2)\big)(y-\mu) \end{aligned}$$

$$+\epsilon e^\top(u,\tau_u)\big(\Sigma^{-1}-\epsilon\Sigma^{-1}\Gamma(u,\tau_u)\Sigma^{-1}+O(\epsilon^2)\big)(y-\mu)$$
$$-\frac{1}{2}\epsilon^2 e^\top(u,\tau_u)\big(\Sigma^{-1}-\epsilon\Sigma^{-1}\Gamma(u,\tau_u)\Sigma^{-1}+O(\epsilon^2)\big)e(u,\tau_u).$$

Since $|\gamma_{ij}(u,\tau_u)|$ and $|e_i(u,\tau_u)|$ are uniformly bounded by 1 w.r.t. τ_u for all $u>U$, we derive that for any $y\in\mathbb{R}^N$,

$$\max_{\tau_u}|d_{u,\tau_u}(y)|\to 0,\ \text{ as } u\to\infty. \tag{2.5.70}$$

By (2.5.69) and (2.5.70), for $y\in\mathbb{R}^N$,

$$\max_{\tau_u}\left|\frac{f_{u,\tau_u(y)}}{f(y)}-1\right|=\max_{\tau_u}\left|e^{d_{u,\tau_u}(y)}\frac{|\Sigma(u,\tau_u)|^{-1/2}}{|\Sigma|^{-1/2}}-1\right|\to 0,\ \text{ as } u\to\infty. \tag{2.5.71}$$

If we could further find an integrable function $g(y)$ on $\mathbb{R}^N$,

$$f(y)\max_{\tau_u}\left|\frac{f_{u,\tau_u}(y)}{f(y)}-1\right|\le g(y), \tag{2.5.72}$$

then (2.5.68) holds by the dominated convergence theorem.

Given a constant C_0, let $A_I:=\{(a_{ij})_{i,j=1}^n\in\mathbb{R}^{N\times N}\mid \max_{i,j}|a_{i,j}|\le C_0\}$, $b_I:=\{(b_i)_{i=1}^n\in\mathbb{R}^N\mid \max_i|b_i|\le C_0\}$. Then there exist constants C_2,C_3, such that

$$|x^\top Ax|\le C_2x^\top x,\quad |b^\top x|\le C_3+x^\top x,\ \ \forall x\in\mathbb{R}^N,\forall A\in A_I,\forall b\in b_I.$$

Hence, there exists a constant $C_4>0$ such that

$$|d_{u,\tau_u}(y)|\le C_4\epsilon(y-\mu)^\top(y-\mu)+C_4\epsilon, \tag{2.5.73}$$

By (2.5.69) and (2.5.73), for small ϵ and large u, there exists a constant K such that

$$\max_{\tau_u} \left| \frac{f_{u,\tau_u(y)}}{f(y)} - 1 \right| \leq K e^{C_4 \epsilon (y-\mu)^\top (y-\mu)} + 1$$

On the other hand, for all $y \in \mathbb{R}^N$,

$$f(y) \leq (2\pi)^{-n/2} |\Sigma|^{-1/2} e^{-\frac{\lambda}{2}(y-\mu)^\top (y-\mu)},$$

where λ is the minimum eigenvalue of Σ. If we choose $\epsilon < \frac{\lambda}{2C_4}$ and define

$$g(y) := (2\pi)^{-n/2} |\Sigma|^{-1/2} e^{-\frac{\lambda}{2}(y-\mu)^\top (y-\mu)} (K e^{C_4 \epsilon (y-\mu)^\top (y-\mu)} + 1),$$

then (2.5.72) holds and hence we have completed the proof. $\square$

Chapter 3

Joint asymptotics of estimating the fractal indices of bivariate Gaussian random processes

Characterizing the dependence structure of multivariate random fields plays a key role in multivariate spatial model setting. Usually, the covariance structure for each component of multivariate processes is highly related to the smoothness of the surface. The estimation of smoothness parameters in univariate model has been studied extensively. Yet, there is few work in the multivariate case. In this chapter, we first give a short review on the increment-based estimator introduced by Kent and Wood [KW97] and apply it to estimating the fractal indices (smoothness parameters) of bivariate Gaussian processes. Then, under the infill asymptotics framework, we investigate the joint asymptotics of the estimators and study how the cross dependence structure would affect the performance of the estimators.

3.1 Introduction

The fractal or Hausdorff dimension of a random process, is a measure of roughness of its sample path. It is an important parameter in geostatistics modeling. Estimating the fractal dimension of a real valued Gaussian and non-Gaussian process has been an attracting problem in the last decades. Hall and Wood [HW93] studied the asymptotic properties of the box-counting estimator for the fractal dimension. The variogram method was introduced by Constantine and Hall [CH94]. Kent and Wood [KW97] developed the increment-based estimators for stationary Gaussian random process, which indicated improved performance under infill asymptotics (that is asymptotic properties of statistical procedures as the sampling points grow dense in a fixed domain [CSY00, Cre93]). Chan and Wood extended the method to Gaussian and a class of non-Gaussian random fields on $\mathbb{R}^2$ (see, e.g., [CW00], [CW04]). Zhu and Stein [ZS02] expanded the work of [CW00] by considering the two-dimensional fractal Brownian surface. We refer to [GŠP12] for further information on this topic.

On the other hand, multivariate (vector-valued) Gaussian random fields have been popular in modeling multivariate spatial datasets (see, e.g., [GKS10]). Usually, the fractal dimension for each component of the multivariate Gaussian random fields varies from each other. It is natural to employ the increment-based methods by Kent and Wood to estimate the fractal dimension for each component. Yet, the joint asymptotic property of the estimators would be non-trivial, since the cross covariance structure might affect the performance of the estimators, that is the covariance among components of multivariate Gaussian random fields. In this work, we study the joint asymptotic properties of estimating fractal indices for bivariate Gaussian random processes under infill asymptotics. The rest of the chapter is organized as follows. We define the bivariate Gaussian random processes in Section 3.2 and introduce the increment-based estimators in Section 3.3. Section 3.4 states the main results on the joint asymptotics of the bivariate estimators. We give an example in Section 3.5. The proofs of our main results are given in Section 3.6.

3.2 The bivariate Gaussian random processes

Let $\{\mathbf{X}(t) \triangleq (X_1(t), X_2(t))^\top, t \in \mathbb{R}\}$ be a bivariate stationary Gaussian random field with mean $\mathbb{E}\mathbf{X}(t) = 0$ and matrix-valued covariance function

$$C(t) = \begin{pmatrix} C_{11}(t) & C_{12}(t) \\ C_{21}(t) & C_{22}(t) \end{pmatrix}$$

where $C_{ij}(t) := \mathbb{E}[X_i(s)X_j(s+t)],\ i = 1, 2$. Further, we assume that

$$\begin{aligned}
C_{11}(t) &= \sigma_1^2 - c_{11}|t|^{\alpha_{11}} + o(|t|^{\alpha_{11}}), \\
C_{22}(t) &= \sigma_2^2 - c_{22}|t|^{\alpha_{22}} + o(|t|^{\alpha_{22}}), \\
C_{12}(t) &= C_{21}(t) = \rho\sigma_1\sigma_2(1 - c_{12}|t|^{\alpha_{12}} + o(|t|^{\alpha_{12}})),
\end{aligned} \tag{3.2.1}$$

with $\alpha_{11}, \alpha_{22} \in (0, 2)$, $\sigma_1, \sigma_2 > 0$, $\rho \in (-1, 1)$ and $c_{11}, c_{22}, c_{12} > 0$. The fractal dimensions for X_1 and X_2 are $2-\alpha_{11}$ and $2-\alpha_{22}$ respectively (see, e.g., [Adl81] Theorem 8.4.1). Hence, we study the estimation and inference of α_{11} and α_{22} instead.

Let F_{11}, F_{22} and F_{12} be the corresponding spectral measure of $C_{11}(\cdot), C_{22}(\cdot)$ and $C_{12}(\cdot)$. By Tauberian Theorem (see, e.g., [Ste99]), we have

$$F_{ij}(x, \infty) \sim C_{ij}(0) - C_{ij}(1/x) \sim |x|^{-\alpha_{ij}},\ i, j = 1, 2.$$

According to Cramer's theorem ([Yag87], [CD09] and [Wac03]), a valid covariance function for $\mathbf{X}(t)$ should satisfy

$$(F_{12}(B))^2 \leq F_{11}(B)F_{22}(B),\ \forall B \in \mathcal{B}(\mathbb{R}).$$

Hence, it is necessary to add the following restriction to $(\alpha_{11}, \alpha_{22}, \alpha_{12})$, i.e.,

$$\frac{\alpha_{11} + \alpha_{22}}{2} \leq \alpha_{12}. \tag{3.2.2}$$

3.3 The increment-based estimators

Assume that X are observed regularly on $[0, 1]$. Specifically, we have n pairs of observations $(X_1(\frac{1}{n}), X_2(\frac{1}{n})), (X_1(\frac{2}{n}), X_2(\frac{2}{n})), ..., (X_1(1), X_2(1))$. Kent and Wood [KW97] introduced the increment-based method to estimate the fractal dimension of a real valued locally self-similar Gaussian process. We apply their methods to estimate the fractal indices for each component of the bivariate Gaussian process (i.e., α_{11}, α_{22}). In Section 3.3.1, we give a review for the definition of the dilated filtered discretized processes and study the asymptotic properties of the covariance of the bivariate dilated filtered discretized processes. In Section 3.3.2, the GLS estimators for the fractal indices $(\alpha_{11}, \alpha_{22})^\top$ are introduced.

3.3.1 The dilated filtered discretized processes

Definition 3.3.1 (Increment of order p). For $J \in \mathbb{Z}^+$ and $p \in \mathbb{Z}^+ \cup 0$, a finite vector $\mathbf{a} = \{a_j\}_{j=-J}^{J}$ is an increment of order p if

$$\sum_{j=-J}^{J} j^r a_j = 0 \text{ for all integer } r \in [0, p] \text{ and } \sum_{j=-J}^{J} j^{p+1} a_j \neq 0. \tag{3.3.1}$$

Definition 3.3.2 (Dilation of a). For an increment $\mathbf{a} = \{a_j\}_{j=-J}^{J}$, the finite vector $\mathbf{a}^u$ is called the dilation of $\mathbf{a}$ for integer $u \geq 1$ if for $-Ju \leq j \leq Ju$,

$$a_j^u = \begin{cases} a_{j'}, & \text{if } j = j'u \\ 0, & \text{otherwise.} \end{cases} \tag{3.3.2}$$

Definition 3.3.3 (**Dilated filtered discretized process**)**.** For $n, m \geq 1$, define the dilated filtered discretized process $Y_{n,i}^u(\cdot)$ by

$$Y_{n,i}^u(j) := n^{\alpha_{ii}/2} \sum_{k=-Ju}^{Ju} a_k^u X_i\left(\frac{j+k}{n}\right), \; i = 1, 2, u = 1, 2, ..., m, \text{ and } j = 1, 2, ..., n. \tag{3.3.3}$$

We give two examples in the following.

- **First-difference increment:** $p = 0$ and $J = 1$ with $a_{-1} = 0, a_0 = -1, a_1 = 1$. Then, the dilated filtered discretized process is

$$Y_{n,i}^u(j) := n^{\frac{\alpha_{ii}}{2}} \left(X_i\left(\frac{j+u}{n}\right) - X_i\left(\frac{j}{n}\right)\right), \tag{3.3.4}$$

  where $i = 1, 2, u = 1, 2, ..., m,$ and $j = 1, 2, ..., n$.

- **Second-difference increment:** $p = 1$ and $J = 1$ with $a_{-1} = 1, a_0 = -2, a_1 = 1$. Then, the dilated filtered discretized process is

$$Y_{n,i}^u(j) := n^{\frac{\alpha_{ii}}{2}} \left(X_i\left(\frac{j-u}{n}\right) - 2X_i\left(\frac{j}{n}\right) + X_i\left(\frac{j+u}{n}\right)\right), \tag{3.3.5}$$

  where $i = 1, 2, u = 1, 2, ..., m,$ and $j = 1, 2, ..., n$.

Next, we consider the covariance of $Y_{n,i}^u, i = 1, 2$. The marginal covariance function (see, e.g., [KW97]) for $Y^u_{n,i}$ is

$$\sigma^{uv}_{n,ii}(h) := \mathbb{E}[Y^u_{n,i}(l)Y^v_{n,i}(l+h)] = n^{\alpha_{ii}} \sum_{j,k} a^u_j a^v_k C_{ii}\Big(\frac{h+k-j}{n}\Big)$$
$$\to -c_{ii} \sum_{j,k} a^u_j a^v_k |h+k-j|^{\alpha_{ii}} \triangleq \sigma^{uv}_{0,ii}(h), \text{ as } n \to \infty. \tag{3.3.6}$$

Especially,

$$Var[Y^u_{n,i}(l)] = \sigma^{uu}_{n,ii}(0) \to \sigma^{uu}_{0,ii}(0) = const. \cdot u^{\alpha_{ii}} \tag{3.3.7}$$

The cross covariance between $Y^u_{n,1}$ and $Y^v_{n,2}$ is given by

$$\sigma^{uv}_{n,12}(h) := \mathbb{E}[Y^u_{n,1}(l)Y^v_{n,2}(l+h)] = n^{\frac{\alpha_{11}+\alpha_{22}}{2}} \sum_{j,k} a^u_j a^v_k C_{12}\Big(\frac{h+k-j}{n}\Big)$$
$$= -c_{12} n^{\frac{\alpha_{11}+\alpha_{22}}{2}} \sum_{j,k} a^u_j a^v_k \Big|\frac{h+k-j}{n}\Big|^{\alpha_{12}} + o(n^{\alpha_{11}/2+\alpha_{22}/2-\alpha_{12}})$$
$$\to \sigma^{uv}_{0,12}(h) \triangleq \begin{cases} 0, & \text{if } \frac{\alpha_{11}+\alpha_{22}}{2} < \alpha_{12} \\ -c_{12} \sum_{j,k} a^u_j a^v_k |h+k-j|^{\alpha_{12}}, & \text{if } \frac{\alpha_{11}+\alpha_{22}}{2} = \alpha_{12}. \end{cases} \tag{3.3.8}$$

Especially,

$$\mathrm{Cov}[Y^u_{n,1}(l), Y^u_{n,2}(l)] = \sigma^{uu}_{n,12}(0) \to \sigma^{uu}_{0,12}(0) = \begin{cases} 0, & \text{if } \frac{\alpha_{11}+\alpha_{22}}{2} < \alpha_{12} \\ const. \cdot u^{\frac{\alpha_{11}+\alpha_{22}}{2}}, & \text{if } \frac{\alpha_{11}+\alpha_{22}}{2} = \alpha_{12}. \end{cases} \tag{3.3.9}$$

Therefore, if $(\alpha_{11}+\alpha_{22})/2<\alpha_{12}$, the covariance of $(Y^u_{n,1}(l),Y^v_{n,2}(l+h))^\top$ satisfies

$$Var\begin{pmatrix} Y^u_{n,1}(l) \\ \\ Y^v_{n,2}(l+h)\end{pmatrix}\to\begin{pmatrix}\sigma^{uv}_{0,11}(h) & 0 \\ \\ 0 & \sigma^{uv}_{0,22}(h)\end{pmatrix},\ \text{as } n\to\infty. \tag{3.3.10}$$

If $(\alpha_{11}+\alpha_{22})/2=\alpha_{12}$ the covariance of $(Y^u_{n,1}(l),Y^v_{n,2}(l+h))^\top$ satisfies

$$Var\begin{pmatrix} Y^u_{n,1}(l) \\ \\ Y^v_{n,2}(l+h)\end{pmatrix}\to\begin{pmatrix}\sigma^{uv}_{0,11}(h) & \sigma^{uv}_{0,12}(h) \\ \\ \sigma^{uv}_{0,12}(h) & \sigma^{uv}_{0,22}(h)\end{pmatrix},\ \text{as } n\to\infty. \tag{3.3.11}$$

3.3.2 The GLS estimators for $(\alpha_{11},\alpha_{22})^\top$

Define

$$Z^u_{n,i}(j):=(Y^u_{n,i}(j))^2,\ j=1,2,...,n, \tag{3.3.12}$$

and

$$\bar{Z}^u_{n,i}:=\frac{1}{n}\sum_{j=1}^{n}Z^u_{n,i}(j), \tag{3.3.13}$$

where $i=1,2,u=1,2,...,m$. By (3.3.7), it is easy to see that

$$\bar{Z}^u_{n,i}\xrightarrow{p}A_iu^{\alpha_{ii}},,i=1,2. \tag{3.3.14}$$

where $A_i,i=1,2$ are constants and hence

$$\log\bar{Z}^u_{n,i}\approx\alpha_{ii}\log u+\log A_i. \tag{3.3.15}$$

Recall the GLS estimator in [KW97]. Let

$$U^{(i)} = (\log \bar{Z}_{n,i}^{1}, ..., \log \bar{Z}_{n,i}^{m})^\top, \; X = (\log 1, \log 2, ..., \log m)^\top \tag{3.3.16}$$

and $\mathbf{1}$ be the m-vector of 1. The generalized least square estimator $\hat{\alpha}_{ii}, i = 1, 2$ is determined by minimizing

$$(U^{(i)} - A_i\mathbf{1} - \alpha_{ii}X)^\top W(U^{(i)} - A_i\mathbf{1} - \alpha_{ii}X), \tag{3.3.17}$$

with respect to α_{ii} and A_i. Hence, we have

$$\hat{\alpha}_{ii} = \frac{(\mathbf{1}^\top W\mathbf{1})(X^\top WY) - (\mathbf{1}^\top WX)(\mathbf{1}^\top WY)}{(\mathbf{1}^\top W\mathbf{1})(X^\top WX) - (\mathbf{1}^\top WX)^2} \tag{3.3.18}$$

- If W is chosen as identity matrix, the GLS estimator is reduced to ordinary least square (abbr. OLS) estimator.
- A good choice of the weighted matrix W in (3.3.17) is the inverse matrix of the covariance of $U^{(i)}$. Let $\Omega^{(i)} = \{\omega_i^{uv}, u, v = 1, ..., m\}$ be the covariance matrix of $U^{(i)}$, which will be specified at the end of Section 3.4.

3.4 Asymptotic properties

In this section, we study the properties of the estimators $(\hat{\alpha}_{11}, \hat{\alpha}_{22})^\top$ under infill asymptotics.

3.4.1 Variance of $\bar{Z}_n$ and asymptotic normality

Let us introduce the notation first. Let

$$\bar{Z}_{n,i} = (\bar{Z}^1_{n,i}, \bar{Z}^2_{n,i}, ..., \bar{Z}^m_{n,i})^\top, i = 1, 2. \tag{3.4.1}$$

and

$$\bar{Z}_n = (\bar{Z}^\top_{n,1}, \bar{Z}^\top_{n,2})^\top. \tag{3.4.2}$$

We apply the method of derivation in Section 3 of [KW97] to the bivariate processes. Recall that $\mathrm{Cov}(Y^u_{n,i}(l), Y^v_{n,j}(l+h)) = \sigma^{uv}_{n,ij}(h), i, j = 1, 2$. Using the fact that, if $(U, V) \sim \mathrm{Normal}(0, 0, 1, 1, \xi)$, then $\mathrm{Cov}(U^2, V^2) = 2\xi^2$, we obtain

$$\mathrm{Cov}(Z^u_{n,i}(l), Z^v_{n,j}(l+h)) = 2(\sigma^{uv}_{n,ij}(h))^2, i, j = 1, 2. \tag{3.4.3}$$

Further,

$$\phi^{uv}_{n,ij} := \mathrm{Cov}(\bar{Z}^u_{n,i}, \bar{Z}^v_{n,j}) = \frac{1}{n} \sum_{h=-n+1}^{n-1} \left(1 - \frac{|h|}{n}\right) \times 2(\sigma^{uv}_{n,ij}(h))^2. \tag{3.4.4}$$

Let $\Phi_{n,ij} = (\phi^{uv}_{n,ij}, u, v = 1, 2, ..., m)$ be the $m \times m$ covariance matrix of $\bar{Z}_{n,i}$ and $\bar{Z}_{n,j}, i, j = 1, 2$. So the covariance matrix of $\bar{Z}_n$ can be written as

$$\Phi_n := \begin{pmatrix} \Phi_{n,11} & \Phi_{n,12} \\ \Phi_{n,12} & \Phi_{n,22} \end{pmatrix}. \tag{3.4.5}$$

In order to study the asymptotic properties of Φ_n, we begin with the asymptotic behavior of $\sigma^{uv}_{n,ij}(h)$ for all $n > |h|$ as $|h| \to \infty$. First of all, it is necessary to make some mild strengthening of Assumption (3.2.1), denoted here by (A_0). For $q \geq 1$, consider the regularity conditions on the qth derivative of $C_{ij}(t)$, say (A_q), for $\forall t \neq 0$,

$$
\begin{aligned}
C^{(q)}_{11}(t) &= -\text{sgn}(t)^q \frac{c_{11}\alpha_{11}!}{(\alpha_{11}-q)!}|t|^{\alpha_{11}-q} + o(|t|^{\alpha_{11}-q}),\\
C^{(q)}_{22}(t) &= -\text{sgn}(t)^q \frac{c_{22}\alpha_{22}!}{(\alpha_{22}-q)!}|t|^{\alpha_{22}-q} + o(|t|^{\alpha_{22}-q}),\\
C^{(q)}_{12}(t) &= C^{(q)}_{21}(t) = -\text{sgn}(t)^q \rho\sigma_1\sigma_2 \frac{c_{12}\alpha_{12}!}{(\alpha_{12}-q)!}|t|^{\alpha_{12}-q} + o(|t|^{\alpha_{12}-q})),
\end{aligned}
\tag{3.4.6}
$$

where $\text{sgn}(t) = 1$ if $t > 0$ and $\text{sgn}(t) = -1$ if $t < 0$.

The theorems below extend Theorem 1 and Theorem 2 in [KW97] to the bivariate case.

Theorem 3.4.1. *If the increment* $\mathbf{a}$ *has order* $p \geq 0$ *and the condition* (A_{2p+2}) *holds, then*

$$
\sigma^{uv}_{n,ii}(h) = O(|h|^{\alpha_{ii}-2p-2}), \text{ as } |h| \to \infty \text{ uniformly for } n > |h|, i = 1, 2, \tag{3.4.7}
$$

and

$$
\begin{aligned}
\sigma^{uv}_{n,12}(h) = n^{\alpha_1/2+\alpha_2/2-\alpha_{12}} O(|h|^{\alpha_{12}-2p-2}) = O(|h|^{\frac{\alpha_{11}+\alpha_{22}}{2}-2p-2}),\\
\text{as } |h| \to \infty \text{ uniformly for } n > |h|.
\end{aligned}
\tag{3.4.8}
$$

By [KW97], we've known using increment with order $p = 1$ will achieve more efficiency. So we'll consider the convergence of variance when $p = 1$. Let $\phi^{uv}_{0,ij} = 2\sum_{h=-\infty}^{\infty}(\sigma^{uv}_{0,ij}(h))^2$,

$\Phi_{0,ij} = (\phi_{0,ij}^{uv}, u, v = 1, 2, ..., m)$, and

$$\Phi_0 := \begin{pmatrix} \Phi_{0,11} & \Phi_{0,12} \\ \Phi_{0,12} & \Phi_{0,22} \end{pmatrix} \tag{3.4.9}$$

Theorem 3.4.2. *If the condition* (A_4) *holds and* $0 < \alpha_{11}, \alpha_{22} < 2$*, then*

$$n\Phi_n \to \Phi_0, \ \textit{as } n \to \infty, \tag{3.4.10}$$

where the entry of Φ_0 *is an absolutely convergent series.*

Theorem 3.4.3. *If the condition* (A_4) *holds and* $0 < \alpha_{11}, \alpha_{22} < 2$*, then*

$$n^{1/2}(\bar{Z}_n - \mathbb{E}[\bar{Z}_n]) \xrightarrow{d} N_{2m}(0, \Phi_0), \ \textit{as } n \to \infty, \tag{3.4.11}$$

3.4.2 Linear estimators of $(\alpha_{11}, \alpha_{22})^\top$

To describe the asymptotic properties of the estimators of $(\alpha_{11}, \alpha_{22})^\top$, it is necessary to specify the remainder term in the assumption (3.2.1). Suppose that, for some $\beta_{11}, \beta_{22}, \beta_{12} > 0$,

$$\begin{aligned}
C_{11}(t) &= \sigma_1^2 - c_{11}|t|^{\alpha_{11}} + O(|t|^{\alpha_{11}+\beta_{11}}), \\
C_{22}(t) &= \sigma_2^2 - c_{22}|t|^{\alpha_{22}} + O(|t|^{\alpha_{22}+\beta_{22}}), \\
C_{12}(t) &= C_{21}(t) = \rho\sigma_1\sigma_2(1 - c_{12}|t|^{\alpha_{12}} + O(|t|^{\alpha_{12}+\beta_{12}})).
\end{aligned} \tag{3.4.12}$$

Here, we will study the asymptotic properties of a more general estimators (see, e.g., [CW00]), that is

$$\hat{\alpha}_{ii} = \sum_{u=1}^{m} L_{u,i} \log \bar{Z}_{n,i}^{u}, \;\; i = 1, 2, \tag{3.4.13}$$

where $L_{u,i}, u = 1, 2, ..., m, i = 1, 2$ are any fixed numbers such that

$$\sum_{u=1}^{m} L_{u,i} = 0 \text{ and } \sum_{u=1}^{m} L_{u,i} \log u = 1. \tag{3.4.14}$$

It is easy to check that the GLS estimator (3.3.18) is an example of the above estimators.

Three theorems below illustrate the asymptotic properties of $(\hat{\alpha}_{11}, \hat{\alpha}_{22})^\top$ by studying the bias, mean square error matrix and asymptotic normality.

Theorem 3.4.4 (Bias)**.** *For the $\hat{\alpha}_{ii}, i = 1, 2$ defined above, we have*

$$\mathbb{E}[\hat{\alpha}_{ii} - \alpha_{ii}] = O(n^{-1}) + O(n^{-\beta_{ii}}), i = 1, 2. \tag{3.4.15}$$

Theorem 3.4.5 (Mean square error matrix)**.** *Let $\hat{\alpha} = (\hat{\alpha}_{11}, \hat{\alpha}_{22})^\top$ and $\alpha = (\alpha_{11}, \alpha_{22})^\top$. If $\alpha_{11} + \alpha_{22} = 2\alpha_{12}$, we have*

$$\mathbb{E}[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^\top] = \begin{pmatrix} O(n^{-1}) + O(n^{-2\beta_{11}}) & O(n^{-1}) + O(n^{-\beta_{11}-\beta_{22}}) \\ O(n^{-1}) + O(n^{-\beta_{11}-\beta_{22}}) & O(n^{-1}) + O(n^{-2\beta_{22}}) \end{pmatrix} \tag{3.4.16}$$

If $\alpha_{11} + \alpha_{22} < 2\alpha_{12}$, we have

$$\mathbb{E}[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^\top]$$

$$= \begin{pmatrix} O(n^{-1}) + O(n^{-2\beta_{11}}) & \begin{matrix} o(n^{-1}) + O(n^{-1-\beta_{11}}) \\ +O(n^{-1-\beta_{22}}) + O(n^{-\beta_{11}-\beta_{22}}) \end{matrix} \\ \begin{matrix} o(n^{-1}) + O(n^{-1-\beta_{11}}) \\ +O(n^{-1-\beta_{22}}) + O(n^{-\beta_{11}-\beta_{22}}) \end{matrix} & O(n^{-1}) + O(n^{-2\beta_{22}}) \end{pmatrix} \tag{3.4.17}$$

Finally, we show the asymptotic normality of the $\hat{\alpha}$. We introduce the notation first. Let

$$T_{n,i}^u = \frac{\bar{Z}_{n,i}^u - \mathbb{E}\bar{Z}_{n,i}^u}{\mathbb{E}\bar{Z}_{n,i}^u},\ T_{n,i} = (T_{n,i}^1, ..., T_{n,i}^m)^\top,$$

$$L_i = (L_{1,i}, ..., L_{m,i})^\top,\ \tilde{L}_i = (L_{1,i}/\sigma_{0,ii}^{11}(0), ..., L_{m,i}/\sigma_{0,ii}^{mm}(0))^\top,\ i = 1, 2.$$

Theorem 3.4.6 (Asymptotic normality)**.** *Assume that $\beta_{11}, \beta_{22} > \frac{1}{2}$, $\sqrt{n}(\hat{\alpha} - \alpha)$ follows the asymptotic properties below.*

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}_{11} - \alpha_{11} \\ \hat{\alpha}_{22} - \alpha_{22} \end{pmatrix} = \begin{pmatrix} \sqrt{n}L_1^\top T_{n,1} + o_p(1) + O(n^{-\beta_{11}+\frac{1}{2}}) \\ \sqrt{n}L_2^\top T_{n,2} + o_p(1) + O(n^{-\beta_{22}+\frac{1}{2}}) \end{pmatrix}, \tag{3.4.18}$$

where

$$\begin{pmatrix} \sqrt{n}L_1^\top T_{n,1} \\ \sqrt{n}L_2^\top T_{n,2} \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \Sigma_\alpha), \tag{3.4.19}$$

with

$$\Sigma_\alpha = \begin{pmatrix} \tilde{L}_1^\top \Phi_{0,11} \tilde{L}_1 & \tilde{L}_1^\top \Phi_{0,12} \tilde{L}_2 \\ \tilde{L}_2^\top \Phi_{0,21} \tilde{L}_1 & \tilde{L}_2^\top \Phi_{0,22} \tilde{L}_2 \end{pmatrix}. \tag{3.4.20}$$

Especially, if $\frac{\alpha_{11}+\alpha_{22}}{2} < \alpha_{12}$, $\Phi_{0,12} = \Phi_{0,21} = \mathbf{0}$ and hence $\sqrt{n}L_1^\top T_{n,1}$ and $\sqrt{n}L_2^\top T_{n,2}$ are *asymptotically independent.*

Remark:

- By Theorem 3.4.6, the covariance matrix of $U^{(i)}$ in Section 3.3.2 can be specified as follows.

$$\omega_i^{uv} = \mathbb{E}[\log \bar{Z}_{n,i}^u \log \bar{Z}_{n,i}^v] \propto n\phi_{n,ii}^{uv}/(\sigma_{n,ii}^{uu}\sigma_{n,ii}^{vv}) \to \phi_{0,ii}^{uv}/(\sigma_{0,ii}^{uu}\sigma_{0,ii}^{vv}). \tag{3.4.21}$$

Hence, ω_i^{uv} is chosen as follows, which is the same as that in Kent and Wood [KW97].

$$\omega_i^{uv} = 2\sum_{h=-n+1}^{n-1}(1-|h|/n)\sigma_{0,ii}^{uv}(h)/\sigma_{0,ii}^{uu}(0)\sigma_{0,ii}^{vv}(0). \tag{3.4.22}$$

3.5 An example: the bivariate Matérn field on $\mathbb{R}$

Recall the definitions of bivariate Matèrn fields and Matèrn correlation functions introduced in Section 2.3. For $m \in \mathbb{Z}$, when $m < \nu < m+1$, we have the following expansion.

$$M(h|\nu, a) = \sum_{j=0}^{m} b_j h^{2j} - b|h|^{2\nu} + o(|t|^{2m+2}),$$

where $b_0, ..., b_m$ are constants and $b = \frac{\Gamma(1-\nu)}{2^{2\nu}\Gamma(1+\nu)}$ (see, e.g., [Ste99], p. 32).

Let $m = 0$. We see that Assumption (3.2.1) is satisfied and the power β_{ij} in Assumption (3.4.12) is $2 - 2\nu_{ij},\ i, j = 1, 2$. Next, we consider the derivatives of the Matérn correlation function. WLOG, assume that $a = 1$ and $M_\nu(h) := M(h|\nu, 1)$. Denote by $\kappa_\nu = \frac{2^{1-\nu}}{\Gamma(\nu)}$. We see that $\kappa_{\nu+1} = (2\nu)^{-1}\kappa_\nu$. Recall that the derivative of the Bessel function of the second kind K_ν satisfies the following recurrence formula.

$$K'_\nu(z) = -K_{\nu+1}(z) + \frac{\nu}{z}K_\nu(z). \tag{3.5.1}$$

Hence,

$$\begin{aligned}
M'_\nu(h) &= \operatorname{sgn}(h)(\kappa_\nu \nu |h|^{\nu-1} K_\nu(|h|) + \kappa_\nu |h|^\nu K'_\nu(|h|)) \\
&= \operatorname{sgn}(h)(2\nu\kappa_\nu |h|^{\nu-1} K_\nu(|h|) - \kappa_\nu |h|^\nu K_{\nu+1}(|h|)) \\
&= 2\nu \cdot \operatorname{sgn}(h)(|h|^{-1}(M_\nu(h) - M_{\nu+1}(h))) \\
&= -2\nu b \cdot \operatorname{sgn}(h)|h|^{2\nu-1} + o(|h|^{2\nu-1}).
\end{aligned} \tag{3.5.2}$$

Similarly,

$$\begin{aligned}
M''_\nu(h) &= (2\nu-1)\operatorname{sgn}(h)|h|^{-1}M'_\nu(h) - 2\nu \cdot \operatorname{sgn}(h)|h|^{-1}M'_{\nu+1}(h) \\
&= -2\nu(2\nu-1)b \cdot \operatorname{sgn}^2(h)|h|^{2\nu-2} + o(|h|^{2\nu-2}), \\
M^{(3)}_\nu(h) &= (2\nu-2)\operatorname{sgn}(h)|h|^{-1}M''_\nu(h) - 2\nu \cdot \operatorname{sgn}(h)|h|^{-1}M''_{\nu+1}(h), \\
&= -2\nu(2\nu-1)(2\nu-2)b \cdot \operatorname{sgn}^3(h)|h|^{2\nu-3} + o(|h|^{2\nu-3}), \\
&\cdots \\
M^{(q)}_\nu(h) &= (2\nu-q+1)\operatorname{sgn}(h)|h|^{-1}M^{(q-1)}_\nu(h) - 2\nu \cdot \operatorname{sgn}(h)|h|^{-1}M^{(q-1)}_{\nu+1}(h) \\
&= -\frac{b(2\nu)!}{(2\nu-q)!}2\nu b \cdot \operatorname{sgn}^q(h)|h|^{2\nu-q} + o(|h|^{2\nu-q}).
\end{aligned} \tag{3.5.3}$$

Hence, Assumption (3.4.6) is satisfied. Now we can apply Theorem 3.4.4 $\sim$ 3.4.6 to bivariate Matérn process.

Theorem 3.5.1 (Bias)**.** *For the nonsmooth bivariate Matérn process with* $0 < \nu_{11}, \nu_{22} < 1$*,* *the bias of* $\hat{\nu}_{ii}$ *is*

$$\mathbb{E}[\hat{\nu}_{ii} - \nu_{ii}] = O(n^{-1}) + O(n^{-(2-\nu_{ii})}), i = 1, 2. \tag{3.5.4}$$

Theorem 3.5.2 (Mean Square Error Matrix)**.** *Let* $\nu = (\nu_{11}, \nu_{22})^\top$ *and* $\hat{\nu} = (\hat{\nu}_{11}, \hat{\nu}_{22})^\top$*. For the nonsmooth bivariate Matérn process with* $0 < \nu_{11}, \nu_{22} < 1$*, if* $\nu_{11} + \nu_{22} = 2\nu_{12}$*, then*

$$\begin{aligned} &\mathbb{E}(\hat{\nu} - \nu)(\hat{\nu} - \nu)^\top \\ &= \begin{pmatrix} O(n^{-1}) + O(n^{-4(1-\nu_{11})}) & O(n^{-1}) + O(n^{-4(1-\nu_{12})}) \\ O(n^{-1}) + O(n^{-4(1-\nu_{12})}) & O(n^{-1}) + O(n^{-4(1-\nu_{22})}) \end{pmatrix} \end{aligned} \tag{3.5.5}$$

If $\nu_{11} + \nu_{22} < 2\nu_{12}$*, we have*

$$\begin{aligned} &\mathbb{E}(\hat{\nu} - \nu)(\hat{\nu} - \nu)^\top \\ &= \begin{pmatrix} O(n^{-1}) + O(n^{-4(1-\nu_{11})}) & o(n^{-1}) + O(n^{-2(2-\nu_{11}-\nu_{22})}) \\ o(n^{-1}) + O(n^{-2(2-\nu_{11}-\nu_{22})}) & O(n^{-1}) + O(n^{-4(1-\nu_{11})}) \end{pmatrix} \end{aligned} \tag{3.5.6}$$

Theorem 3.5.3 (Asymptotic normality)**.** *For the nonsmooth bivariate Matérn process with* $0 < \nu_{11}, \nu_{22} < \frac{3}{4}$,

$$\sqrt{n} \begin{pmatrix} \hat{\nu}_{11} - \nu_{11} \\ \hat{\nu}_{22} - \nu_{22} \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \Sigma_\nu), \tag{3.5.7}$$

where

$$\Sigma_\nu = \left(\begin{array}{cc} \tilde{L}_1^\top \Phi_{0,11} \tilde{L}_1 & \tilde{L}_1^\top \Phi_{0,12} \tilde{L}_2 \\ \tilde{L}_2^\top \Phi_{0,21} \tilde{L}_1 & \tilde{L}_2^\top \Phi_{0,22} \tilde{L}_2 \end{array} \right), \tag{3.5.8}$$

and $\tilde{L}_i = (L_{1,i}/\sigma_{0,ii}^{11}(0), ..., L_{m,i}/\sigma_{0,ii}^{mm}(0))^\top, i = 1, 2$. *Especially, if* $\frac{\nu_{11}+\nu_{22}}{2} < \nu_{12}$, $\Phi_{0,12} = \Phi_{0,21} = \mathbf{0}$ *and hence* $\hat{\nu}_{11}$ *and* $\hat{\nu}_{22}$ *are asymptotically independent.*

3.6 Proof of the main results

Proof of Theorem 3.4.1. (3.4.7) comes directly from the proof of Theorem 1 in [KW97]. We are going to prove (3.4.8). First of all, expand $C_{12}(\frac{h+k-j}{n})$ in a Taylor series about h/n to the $(2p+2)$th order to obtain

$$\begin{aligned} \sigma_{n,12}^{uv}(h) &= n^{\frac{\alpha_{11}+\alpha_{22}}{2}} \sum_{j,k} a_j^u a_k^v C_{12}\left(\frac{h+k-j}{n}\right) \\ &= n^{\frac{\alpha_{11}+\alpha_{22}}{2}} \sum_{r=0}^{2p+1} \sum_{j,k} a_j^u a_k^v \frac{(k-j)^r}{r! n^r} C_{12}^{(r)}\left(\frac{h}{n}\right) \\ &+ n^{\frac{\alpha_{11}+\alpha_{22}}{2}} \sum_{j,k} a_j^u a_k^v \frac{(k-j)^{2p+2}}{(2p+2)! n^{2p+2}} C_{12}^{(2p+2)}\left(\frac{h_{kj}^*}{n}\right) \\ &= n^{\frac{\alpha_{11}+\alpha_{22}}{2}} \sum_{j,k} a_j^u a_k^v \frac{(k-j)^{2p+2}}{(2p+2)! n^{2p+2}} C_{12}^{(2p+2)}\left(\frac{h_{kj}^*}{n}\right), \end{aligned} \tag{3.6.1}$$

where h_{kj}^* lies between h and $h+k-j$. Since $|k-j| \le (u+v)J \le 2mJ$, $h_{kj}^* \le 2|h|$ for all $|h| \ge 2mJ$. Combining the condition (A_{2p+2}) satisfied, for all $|h| \ge 2mJ$ and all $n > |h|$ we have

$$|\sigma_{n,12}^{uv}(h)| \le const. |h|^{\alpha_{12}-2p-2} \cdot n^{\frac{\alpha_{11}+\alpha_{22}}{2}-\alpha_{12}} \le const. |h|^{\frac{\alpha_{11}+\alpha_{22}}{2}-2p-2}.$$

□

Proof of Theorem 3.4.2. The proof is very similar as that in [KW97]. Let

$$d_{n,ij}^{uv}(h) := \begin{cases} \left(1 - \frac{|h|}{n}\right)(\sigma_{n,ij}^{uv}(h))^2, & |h| < n \\ 0, & \text{otherwise.} \end{cases} \tag{3.6.2}$$

By (3.3.6) and (3.3.8), we have $d_{n,ij}^{uv}(h) \to \sigma_{0,ij}^{uv}(h)$ as $n \to \infty$. By Theorem 3.4.1 and $p = 1$, we know

$$d_{n,ij}^{uv}(h) \leq const.|h|^{\alpha_{ii}+\alpha_{jj}-8}.$$

Since $\alpha_{ii} + \alpha_{jj} - 8 < -4$, $\sum_{h=-\infty}^{\infty} d_{n,ij}^{uv}(h)$ is bounded by a summable series. Therefore, (3.4.10) can be concluded by dominated convergence theorem. □

Proof of Theorem 3.4.3. By the Cramér-Wold theorem, it is equivalent to prove that for $\forall \gamma = (\gamma_{1,1}, ..., \gamma_{m,1}, \gamma_{1,2}, ..., \gamma_{m,2})^\top \in \mathbb{R}^{2m}$,

$$n^{1/2}\gamma^\top(\bar{Z}_n - \mathbb{E}[\bar{Z}_n]) \xrightarrow{d} N(0, \gamma^\top \Phi_0 \gamma), \text{ as } n \to \infty, \tag{3.6.3}$$

First, we introduce the notation. Let $\gamma_i := (\gamma_{1,i}, ..., \gamma_{m,i})^\top, i = 1, 2$ and

$$\Gamma_n = diag(\underbrace{\gamma_1^\top, \gamma_1^\top, ..., \gamma_1^\top}_{n \text{ times}}, \underbrace{\gamma_2^\top, \gamma_2^\top, ..., \gamma_2^\top}_{n \text{ times}})^\top. \tag{3.6.4}$$

So Γ_n is a $2mn \times 2mn$ matrix including n copies of γ_1 and γ_2 on the diagonal. Let

$$Y_{n,i}(j) := (Y_{n,i}^1(j), Y_{n,i}^2(j), ..., Y_{n,i}^m(j))^\top, i = 1, 2, j = 1, 2, ..., n, \tag{3.6.5}$$

and

$$W_n = (Y_{n,1}^{\top}(1), Y_{n,1}^{\top}(2), ..., Y_{n,1}^{\top}(n), Y_{n,2}^{\top}(1), Y_{n,2}^{\top}(2), ..., Y_{n,2}^{\top}(n))^{\top}, \tag{3.6.6}$$

where W_n is a $2mn$ dimensional vector.

Then, we have

$$S_n \triangleq n^{1/2}\gamma^{\top}(\bar{Z}_n - \mathbb{E}[\bar{Z}_n]) = n^{-1/2}(W_n^{\top}\Gamma_n W_n - \mathbb{E}[W_n^{\top}\Gamma_n W_n]). \tag{3.6.7}$$

Let

$$V_n = \mathbb{E}[W_n W_n^{\top}], \tag{3.6.8}$$

be the covariance matrix of W_n. For $1 \leq i_1, i_2 \leq 2, 1 \leq j_1, j_2 \leq n, 1 \leq k_1, k_2 \leq m$, let

$$l_1 = (i_1 - 1)mn + (j_1 - 1)m + k_1,$$

$$l_2 = (i_2 - 1)mn + (j_2 - 1)m + k_2,$$

So the (l_1, l_2) entry of W_n is

$$V_n(l_1, l_2) = \mathbb{E}[Y_{n,i_1}^{k_1}(j_1)Y_{n,i_2}^{k_2}(j_2)] = \sigma_{n,i_1 i_2}^{k_1 k_2}(j_2 - j_1). \tag{3.6.9}$$

Let $\widetilde{W_n} = V_n^{-\frac{1}{2}}W_n$ and $\Lambda_n = 2n^{-\frac{1}{2}}V_n^{\frac{1}{2}}\Gamma V_n^{\frac{1}{2}}$. Then we have

$$S_n = \frac{1}{2}(\widetilde{W_n}^{\top}\Lambda_n\widetilde{W_n} - \mathbb{E}[\widetilde{W_n}^{\top}\Lambda_n\widetilde{W_n}]) \tag{3.6.10}$$

It is easy to see $\widetilde{W}_n \sim N_{2mn}(0, I)$ where I is the identity matrix. There exists an orthogonal matrix Q such that $Q^\top \Lambda_n Q$ is a diagonal matrix whose diagonal entries are eigenvalues of Λ_n, denoted by $\lambda_{n,j}, j = 1, 2, ..., 2mn$. Also, $U \triangleq Q\widetilde{W}_n \sim N_{2mn}(0, I)$. Therefore, for $\forall \theta < \min_{1 \le j \le 2mn} \lambda_{n,j}^{-1}$, the cumulant generating function of $\frac{1}{2}\widetilde{W}_n^\top \Lambda_n \widetilde{W}_n$ is given by

$$\log \mathbb{E}\big[e^{\frac{\theta}{2}\widetilde{W}_n^\top \Lambda_n \widetilde{W}_n}\big] = \log \mathbb{E}\big[e^{\frac{\theta}{2}U^\top (Q^\top \Lambda_n Q)U}\big] = \log \mathbb{E}\big[e^{\frac{\theta}{2}\sum_{j=1}^{2mn} \lambda_{n,j} U_j^2}\big]$$

$$= -\frac{1}{2}\sum_{j=1}^{2mn} \log(1 - \theta\lambda_{n,j}). \tag{3.6.11}$$

So the cumulant generating function of S_n is given by

$$k_n(\theta) \triangleq \log \mathbb{E}[e^{\theta S_n}] = -\frac{1}{2}\sum_{j=1}^{2mn}(\log(1 - \theta\lambda_{n,j}) + \theta\lambda_{n,j}) \tag{3.6.12}$$

As the proof of Theorem 3.2 in [KW95], it is sufficient to prove that

$$tr(\Lambda^4) = \sum_{j=1}^{2mn} \lambda_{n,j}^4 \to 0, \text{ as } n \to \infty. \tag{3.6.13}$$

First of all, let's prove why (3.6.13) ensure the asymptotic normality of S_n. The argument is very similar as that in [KW95]). Applying Taylor's expansion to $\log(1 - \theta\lambda_{n,j})$ at $\theta = 0$, we obtain

$$k_n(\theta) = \frac{\theta^2}{4}\sum_{j=1}^{2mn} \lambda_{n,j}^2 + \frac{\theta^3}{6}\sum_{j=1}^{2mn} \lambda_{n,j}^3 + \frac{\theta^4}{8}\sum_{j=1}^{2mn}(1 - \theta_{n,j}\lambda_{n,j})^{-4}\lambda_{n,j}^4, \tag{3.6.14}$$

where $\theta_{n,j}$ is between 0 and θ. Let's consider the term $\frac{1}{2}\sum_{j=1}^{2mn} \lambda_{n,j}^2$. It follows from (3.6.9) that

$$
\begin{aligned}
&\frac{1}{2}\sum_{j=1}^{2mn}\lambda_{n,j}^2=\frac{1}{2}tr(\Lambda_n^2)=\frac{2}{n}tr((V_n\Gamma_n)^2)\\
=&\frac{2}{n}\sum_{l_1}^{2mn}\sum_{l_2}^{2mn}(V_n\Gamma_n)(l_1,l_2)(V_n\Gamma_n)(l_2,l_1)\\
=&\frac{2}{n}\sum_{l_1}^{2mn}\sum_{l_2}^{2mn}V_n(l_1,l_2)\Gamma_n(l_2,l_2)V_n(l_2,l_1)\Gamma(l_1,l_1)\\
=&\frac{2}{n}\sum_{i_1,i_2=1}^{2}\sum_{k_1,k_2=1}^{m}\sum_{j_1,j_2=1}^{n}\sigma_{n,i_1i_2}^{k_1k_2}(j_2-j_1)\gamma_{k_2,i_2}\sigma_{n,i_2i_1}^{k_2k_1}(j_1-j_2)\gamma_{k_1,i_1}\\
=&\frac{2}{n}\sum_{i_1,i_2=1}^{2}\sum_{k_1,k_2=1}^{m}\sum_{j_1,j_2=1}^{n}\gamma_{k_1,i_1}\gamma_{k_2,i_2}\big(\sigma_{n,i_1i_2}^{k_1k_2}(j_2-j_1)\big)^2,
\end{aligned}
\tag{3.6.15}
$$

where

$$l_1=(i_1-1)mn+(j_1-1)m+k_1, 1\le i_1\le 2, 1\le j_1\le n, 1\le k_1\le m,$$

$$l_2=(i_2-1)mn+(j_2-1)m+k_2, 1\le i_2\le 2, 1\le j_2\le n, 1\le k_2\le m.$$

On the other hand,

$$
\begin{aligned}
&\gamma^\top\Phi_n\gamma=(\gamma_1^\top,\gamma_2^\top)\begin{pmatrix}\Phi_{n,11} & \Phi_{n,12}\\ \Phi_{n,12} & \Phi_{n,22}\end{pmatrix}\begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix}\\
=&\sum_{i_1,i_2=1}^{2}\sum_{k_1,k_2=1}^{m}\gamma_{k_1,i_1}\gamma_{k2,i_2}\phi_{n,i_1i_2}^{k_1k_2}\\
=&\frac{1}{n^2}\sum_{i_1,i_2=1}^{2}\sum_{k_1,k_2=1}^{m}\sum_{j_1,j_2=1}^{n}\gamma_{k_1,i_1}\gamma_{k_2,i_2}\mathbb{E}[Z_{n,i}^uZ_{n,j}^v]
\end{aligned}
$$

$$=\frac{2}{n^2}\sum_{i_1,i_2=1}^{2}\sum_{k_1,k_2=1}^{m}\sum_{j_1,j_2=1}^{n}\gamma_{k_1,i_1}\gamma_{k_2,i_2}\big(\sigma_{n,i_1i_2}^{k_1k_2}(j_2-j_1)\big)^2. \tag{3.6.16}$$

Hence, by (3.6.15),(3.6.16) and Theorem 3.4.2, we have

$$\frac{1}{2}\sum_{j=1}^{2mn}\lambda_{n,j}^2=\gamma^\top(n\Phi_n)\gamma\to\gamma^\top\Phi_0\gamma,\ \text{ as } n\to\infty. \tag{3.6.17}$$

Next, let's consider the second term in (3.6.14). By (3.6.13), we have

$$\max_{1\le j\le 2mn}|\lambda_{n,j}|\le\left(\sum_{j=1}^{2mn}\lambda_{n,j}^4\right)^{\frac{1}{4}}\to 0,\ \text{ as } n\to\infty, \tag{3.6.18}$$

which implies

$$\left|\sum_{j=1}^{2mn}\lambda_{n,j}^3\right|\le\max_{1\le j\le 2mn}|\lambda_{n,j}|\sum_{j=1}^{2mn}\lambda_{n,j}^2\to 0,\ \text{ as } n\to\infty. \tag{3.6.19}$$

Finally, let's consider the third term in (3.6.14). By (3.6.18), we know $\delta:=\sup_{n\ge1}\max_{1\le j\le 2mn}|\lambda_{n,j}|$ is positive and finite. If we restrict attention to $|\theta|\le\frac{1}{2}\delta^{-1}$, we have $(1-\theta_{n,j}\lambda_{n,j})^{-4}\le 16$ and hence for $\theta\in(-\frac{1}{2}\delta^{-1},\frac{1}{2}\delta^{-1})$,

$$\sum_{j=1}^{2mn}(1-\theta_{n,j}\lambda_{n,j})^{-4})\lambda_{n,j}^4\to 0,\ \text{ as } n\to\infty. \tag{3.6.20}$$

Therefore, by (3.6.17),(3.6.19) and (3.6.20), for $\forall\theta\in(-\frac{1}{2}\delta^{-1},\frac{1}{2}\delta^{-1})$, we have

$$k_n(\theta)\to\frac{\theta^2}{2}\gamma^\top\Phi_0\gamma, \tag{3.6.21}$$

which is sufficient to prove

$$S_n := n^{1/2}\gamma^\top(\bar{Z}_n - \mathbb{E}[\bar{Z}_n]) \xrightarrow{d} N(0, \gamma^\top \Phi_0 \gamma), \text{ as } n \to \infty,$$

Now we only need to prove (3.6.13).

$$\begin{aligned}
&tr(\Lambda_n^4) = \frac{16}{n^2} tr((V_n\Gamma_n)^4) \\
=&\frac{16}{n^2} \sum_{l_1,l_2,...,l_4=1}^{2mn} (V_n\Gamma_n)(l_1,l_2)(V_n\Gamma_n)(l_2,l_3)(V_n\Gamma_n)(l_3,l_4)(V_n\Gamma_n)(l_4,l_1) \\
=&\frac{16}{n^2} \sum_{i_1,...,i_4=1}^{2} \sum_{k_1,...,k_4=1}^{m} \gamma_{k_1,i_1}\gamma_{k_2,i_2}\gamma_{k_3,i_3}\gamma_{k_4,i_4} \\
&\cdot \sum_{j_1,...,j_4=1}^{n} \sigma_{n,i_1i_2}^{k_1k_2}(j_2-j_1)\sigma_{n,i_2i_3}^{k_2k_3}(j_3-j_2)\sigma_{n,i_3i_4}^{k_3k_4}(j_4-j_3)\sigma_{n,i_4i_1}^{k_4k_1}(j_1-j_4),
\end{aligned} \tag{3.6.22}$$

where $l_r = (i_r-1)mn + (j_r-1)m + k_r, r = 1, ..., 4$. Let

$$\begin{aligned}
&\Delta_n(k_1, ..., k_4, i_1, ..., i_4) \\
:=& \sum_{j_1,...,j_4=1}^{n} \sigma_{n,i_1i_2}^{k_1k_2}(j_2-j_1)\sigma_{n,i_2i_3}^{k_2k_3}(j_3-j_2)\sigma_{n,i_3i_4}^{k_3k_4}(j_4-j_3)\sigma_{n,i_4i_1}^{k_4k_1}(j_4-j_1) \\
=& \sum_{j_1,...,j_4=1}^{n} \sigma_{n,i_1i_2}^{k_1k_2}(h_1)\sigma_{n,i_2i_3}^{k_2k_3}(h_2)\sigma_{n,i_3i_4}^{k_3k_4}(h_3)\sigma_{n,i_4i_1}^{k_4k_1}(h_1+h_2+h_3),
\end{aligned} \tag{3.6.23}$$

where $h_i = j_{i+1} - j_i, i = 1, 2, 3$.

Given h_1, h_2 and h_3 fixed, the cardinality of the set

$$\#\{(j_1, j_2, ..., j_4) \mid 1 \le j_1, ..., j_4 \le n\} \le n. \tag{3.6.24}$$

Hence,

$$
\begin{aligned}
&|\Delta_n(k_1, ..., k_4, i_1, ..., i_4)| \\
\leq & n \sum_{|h_1|,|h_2|,|h_3| \leq n-1} |\sigma_{n,i_1 i_2}^{k_1 k_2}(h_1)\sigma_{n,i_2 i_3}^{k_2 k_3}(h_2)\sigma_{n,i_3 i_4}^{k_3 k_4}(h_3)\sigma_{n,i_4 i_1}^{k_4 k_1}(h_1+h_2+h_3)| \qquad (3.6.25)
\end{aligned}
$$

Further, by Theorem 3.4.1, we have

$$
\begin{aligned}
&|\Delta_n(k_1, ..., k_4, i_1, ..., i_4)| \\
\leq & const.n \prod_{r=1}^{3} \sum_{h_r=-n+1}^{n-1} h_r^{\frac{\alpha_{i_r i_r}}{2}+\frac{\alpha_{i_{r+1} i_{r+1}}}{2}-4} \\
\leq & const.n \prod_{r=1}^{3} \sum_{h_r=-\infty}^{\infty} h_r^{\frac{\alpha_{i_r i_r}}{2}+\frac{\alpha_{i_{r+1} i_{r+1}}}{2}-4} \\
= & O(n). \qquad (3.6.26)
\end{aligned}
$$

The last equality holds since $\frac{\alpha_{i_r i_r}}{2}+\frac{\alpha_{i_{r+1} i_{r+1}}}{2}-4 < -2$. Therefore, by (3.6.22) and (3.6.26), we have

$$tr(\Lambda_n^4) = O(n^{-1}) \to 0, \text{ as } n \to \infty.$$

$\square$

Proof of Theorem 3.4.4. Recall that

$$T_{n,i}^u = \frac{\bar{Z}_{n,i}^u - \mathbb{E}\bar{Z}_{n,i}^u}{\mathbb{E}\bar{Z}_{n,i}^u}. \qquad (3.6.27)$$

If

$$|T_{n,i}^u| \leq \xi \leq \frac{1}{2}, \tag{3.6.28}$$

by Taylor expansion, we have

$$\log(1 + T_{n,i}^u) = T_{n,i}^u - \frac{1}{2}(T_{n,i}^u)^2 + R_{n,i}^u, \tag{3.6.29}$$

where $|R_{n,i}^u| \leq \xi (T_{n,i}^u)^2$. Hence,

$$\mathbb{E}[\log(1 + T_{n,i}^u)] = \mathbb{E}[T_{n,i}^u - \frac{1}{2}(T_{n,i}^u)^2 + R_{n,i}^u] = -\frac{1}{2}\mathbb{E}(T_{n,i}^u)^2 + ER_{n,i}^u, \tag{3.6.30}$$

and

$$\mathbb{E}[|R_{n,i}^u|, |T_{n,i}^u| \leq \xi] \leq \xi \mathbb{E}[(T_{n,i}^u)^2, |T_{n,i}^u| \leq \xi] = O(n^{-1}). \tag{3.6.31}$$

By Lemma 3.6.1, we obtain

$$\sum_{u=1}^m L_{u,i} \mathbb{E}[\log(1 + T_{n,i}^u), |T_{n,i}^u| > \xi] = o(n^{-1}), \tag{3.6.32}$$

and hence

$$\begin{aligned}
&\sum_{u=1}^m L_{u,i} \mathbb{E}[|R_{n,i}^u|, |T_{n,i}^u| > \xi] \\
&\leq \sum_{u=1}^m L_{u,i} \mathbb{E}[|\log(1 + T_{n,i}^u| + |T_{n,i}^u| + \frac{1}{2}(T_{n,i}^u)^2, \, |T_{n,i}^u| > \xi] \\
&= O(n^{-1}).
\end{aligned} \tag{3.6.33}$$

Therefore,

$$\sum_{u=1}^{m} L_{u,i}\mathbb{E}[\log(1+T_{n,i}^{u})] = O(n^{-1}). \tag{3.6.34}$$

On the other hand, by assumption (3.4.12), we have

$$\mathbb{E}\bar{Z}_{n,i}^{u} = \sigma_{n,ii}^{uu}(0) = const.u^{\alpha_{ii}}(1+O(n^{-\beta_{ii}})), \tag{3.6.35}$$

and hence

$$\sum_{u=1}^{m} L_{u,i}\log\mathbb{E}\bar{Z}_{n,i}^{u} = \alpha_{ii}\log u + O(n^{-\beta_{ii}}) \tag{3.6.36}$$

by (3.4.14).

Therefore, by (3.6.34) and (3.6.36),

$$\mathbb{E}[\hat{\alpha}_{ii} - \alpha_{ii}] = \mathbb{E}\bigg[\sum_{u=1}^{m} L_{u,i}(\log\bar{Z}_{n,i}^{u} - \alpha_{ii}\log u)\bigg] = O(n^{-1}) + O(n^{-\beta_{ii}}). \tag{3.6.37}$$

$\square$

Proof of Theorem 3.4.5.

$$\begin{aligned}
&\mathbb{E}(\hat{\alpha}_{ii} - \alpha_{ii})^2 \\
=&\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,i}L_{v,i}\mathbb{E}(\log\bar{Z}_{n,i}^{u} - \alpha_{ii}\log u)(\log\bar{Z}_{n,i}^{v} - \alpha_{ii}\log v) \\
=&\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,i}L_{v,i}\mathbb{E}(\log(1+T_{n,i}^{u}) + \log\mathbb{E}\bar{Z}_{n,i}^{u} - \alpha_{ii}\log u) \\
&\times(\log(1+T_{n,i}^{v}) + \log\mathbb{E}\bar{Z}_{n,i}^{v} - \alpha_{ii}\log v)
\end{aligned}$$

$$
\begin{aligned}
=&\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,i}L_{v,i}\mathbb{E}[\log(1+T_{n,i}^u)\log(1+T_{n,i}^v)]\\
&+\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,i}L_{v,i}\mathbb{E}\log(1+T_{n,i}^u)(\log\mathbb{E}\bar{Z}_{n,i}^v-\alpha_{ii}\log v)\\
&+\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,i}L_{v,i}(\log\mathbb{E}\bar{Z}_{n,i}^u-\alpha_{ii}\log u)\mathbb{E}\log(1+T_{n,i}^v)\\
&+\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,i}L_{v,i}(\log\mathbb{E}\bar{Z}_{n,i}^u-\alpha_{ii}\log u)(\log\mathbb{E}\bar{Z}_{n,i}^v-\alpha_{ii}\log v)
\end{aligned}
\tag{3.6.38}
$$

By (3.6.34) and (3.6.36), we have

$$
\begin{aligned}
&\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,i}L_{v,i}\mathbb{E}\log(1+T_{n,i}^u)(\log\mathbb{E}\bar{Z}_{n,i}^v-\alpha_{ii}\log v)=O(n^{-1-\beta_{ii}})\\
&\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,i}L_{v,i}(\log\mathbb{E}\bar{Z}_{n,i}^u-\alpha_{ii}\log u)\mathbb{E}\log(1+T_{n,i}^v)=O(n^{-1-\beta_{ii}})\\
&\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,i}L_{v,i}(\log\mathbb{E}\bar{Z}_{n,i}^u-\alpha_{ii}\log u)(\log\mathbb{E}\bar{Z}_{n,i}^v-\alpha_{ii}\log v)=O(n^{-2\beta_{ii}}).
\end{aligned}
\tag{3.6.39}
$$

Next, we study the first term in the right hand side of (3.6.38). First,

$$
\begin{aligned}
&\mathbb{E}[\log(1+T_{n,i}^u)\log(1+T_{n,i}^v), |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi]\\
=&\mathbb{E}\big[(T_{n,i}^u-1/2(T_{n,i}^u)^2+R_{n,i}^u)(T_{n,i}^v-1/2(T_{n,i}^v)^2+R_{n,i}^v), |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi\big]\\
=&\mathbb{E}\big[T_{n,i}^uT_{n,i}^v, |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi\big]-\frac{1}{2}\mathbb{E}\big[T_{n,i}^u(T_{n,i}^v)^2, |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi\big]\\
&+\mathbb{E}\big[T_{n,i}^uR_{n,i}^v, |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi\big]-\frac{1}{2}\mathbb{E}\big[(T_{n,i}^u)^2T_{n,i}^v, |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi\big]\\
&+\frac{1}{4}\mathbb{E}\big[(T_{n,i}^u)^2(T_{n,i}^v)^2, |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi\big]-\frac{1}{2}\mathbb{E}\big[(T_{n,i}^u)^2R_{n,i}^v, |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi\big]\\
&+\mathbb{E}\big[R_{n,i}^uT_{n,i}^v, |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi\big]-\frac{1}{2}\mathbb{E}\big[R_{n,i}^u(T_{n,i}^v)^2, |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi\big]\\
&+\mathbb{E}\big[R_{n,i}^uR_{n,i}^v, |T_{n,i}^u|<\xi, |T_{n,i}^v|<\xi\big]
\end{aligned}
\tag{3.6.40}
$$

By Theorem 3.4.2, we have

$$\mathbb{E}[T_{n,i}^u T_{n,i}^v] = O(n^{-1}). \tag{3.6.41}$$

Hence, we have

$$\begin{aligned}
&\mathbb{E}[T_{n,i}^u (T_{n,i}^v)^2] = o(n^{-1}), \mathbb{E}[(T_{n,i}^u)^2 (T_{n,i}^v)^2] = o(n^{-1}) \\
&\mathbb{E}\big[R_{n,i}^u R_{n,i}^v, |T_{n,i}^u| < \xi, |T_{n,i}^v| < \xi\big] = o(n^{-1}), \\
&\mathbb{E}\big[R_{n,i}^u T_{n,i}^v, |T_{n,i}^u| < \xi, |T_{n,i}^v| < \xi\big] = o(n^{-1}), \\
&\mathbb{E}\big[R_{n,i}^u (T_{n,i}^v)^2, |T_{n,i}^u| < \xi, |T_{n,i}^v| < \xi\big] = o(n^{-1})
\end{aligned} \tag{3.6.42}$$

Therefore,

$$\mathbb{E}[\log(1 + T_{n,i}^u) \log(1 + T_{n,i}^v), |T_{n,i}^u| < \xi, |T_{n,i}^v| < \xi] = O(n^{-1}). \tag{3.6.43}$$

Second, by Lemma 3.6.1,

$$\begin{aligned}
&\mathbb{E}[\log(1 + T_{n,i}^u) \log(1 + T_{n,i}^v), |T_{n,i}^u| > \xi, |T_{n,i}^v| < \xi] \\
&\leq |\log(1 - \xi)| \mathbb{E}[|\log(1 + T_{n,i}^u)|, |T_{n,i}^u| > \xi] \\
&= o(n^{-1}).
\end{aligned} \tag{3.6.44}$$

Similarly,

$$\mathbb{E}[\log(1 + T_{n,i}^u) \log(1 + T_{n,i}^v), |T_{n,i}^u| < \xi, |T_{n,i}^v| > \xi]$$

$$
\begin{aligned}
&\leq |\log(1-\xi)|\mathbb{E}[|\log(1+T^v_{n,i})|, |T^v_{n,i}| > \xi] \\
&= o(n^{-1}).
\end{aligned} \tag{3.6.45}
$$

Finally, by Lemma 3.6.1,

$$
\begin{aligned}
&\mathbb{E}[\log(1+T^u_{n,i})\log(1+T^v_{n,i}), |T^u_{n,i}| > \xi,\, |T^v_{n,i}| > \xi] \\
&\leq (\mathbb{E}[\log^2(1+T^u_{n,i}), |T^u_{n,i}| > \xi])^{1/2}(\mathbb{E}[\log(1+T^v_{n,i}), |T^v_{n,i}| > \xi])^{1/2} \\
&= o(n^{-1}).
\end{aligned} \tag{3.6.46}
$$

Therefore,

$$
\mathbb{E}[\log(1+T^u_{n,i})\log(1+T^v_{n,i})] = O(n^{-1}). \tag{3.6.47}
$$

By (3.6.39) and (3.6.47), we have

$$
\mathbb{E}(\hat{\alpha}_{ii} - \alpha_{ii})^2 = O(n^{-1}) + O(n^{-2\beta_{ii}}), i = 1, 2 \tag{3.6.48}
$$

Next, we study the cross term $\mathbb{E}(\hat{\alpha}_{11} - \alpha_{11})(\hat{\alpha}_{22} - \alpha_{22})$. Similarly as $\mathbb{E}(\hat{\alpha}_{ii} - \alpha_{ii})^2$,

$$
\begin{aligned}
&\mathbb{E}(\hat{\alpha}_{11} - \alpha_{11})(\hat{\alpha}_{22} - \alpha_{22}) \\
&= \sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,1}L_{v,2}\mathbb{E}(\log \bar{Z}^u_{n,1} - \alpha_{11}\log u)(\log \bar{Z}^v_{n,2} - \alpha_{22}\log v) \\
&= \sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,1}L_{v,2}\mathbb{E}[\log(1+T^u_{n,1})\log(1+T^v_{n,2})] \\
&\quad + \sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,1}L_{v,2}\mathbb{E}\log(1+T^u_{n,1})(\log \mathbb{E}\bar{Z}^v_{n,2} - \alpha_{22}\log v)
\end{aligned}
$$

$$+\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,1}L_{v,2}(\log \mathbb{E}\bar{Z}_{n,1}^{u} - \alpha_{11}\log u)\mathbb{E}\log(1+T_{n,2}^{v})$$
$$+\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,1}L_{v,2}(\log \mathbb{E}\bar{Z}_{n,1}^{u} - \alpha_{11}\log u)(\log \mathbb{E}\bar{Z}_{n,2}^{v} - \alpha_{22}\log v). \tag{3.6.49}$$

So we only need to find the order of $\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,1}L_{v,2}\mathbb{E}[\log(1+T_{n,1}^{u})\log(1+T_{n,2}^{v})]$.

Case 1**:** If $\frac{\alpha_{11}+\alpha_{22}}{2} = \alpha_{12}$, by Theorem 3.4.2 and (3.3.11),

$$\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,1}L_{v,2}\mathbb{E}[\log(1+T_{n,1}^{u})\log(1+T_{n,2}^{v})] = O(n^{-1}), \tag{3.6.50}$$

and hence

$$\mathbb{E}(\hat{\alpha}_{11}-\alpha_{11})(\hat{\alpha}_{22}-\alpha_{22}) = O(n^{-1}) + O(n^{-\beta_{11}-\beta_{22}}), i = 1, 2 \tag{3.6.51}$$

Case 2: If $\frac{\alpha_{11}+\alpha_{22}}{2} < \alpha_{12}$, by Theorem 3.4.2 and (3.3.10),

$$\sum_{u=1}^{m}\sum_{v=1}^{m} L_{u,1}L_{v,2}\mathbb{E}[\log(1+T_{n,1}^{u})\log(1+T_{n,2}^{v})] = o(n^{-1}), \tag{3.6.52}$$

and hence

$$\begin{aligned}&\mathbb{E}(\hat{\alpha}_{11}-\alpha_{11})(\hat{\alpha}_{22}-\alpha_{22})\\ &= o(n^{-1}) + O(n^{-1-\beta_{11}}) + O(n^{-1-\beta_{22}}) + O(n^{-\beta_{11}-\beta_{22}}).\end{aligned} \tag{3.6.53}$$

□

Proof of Theorem 3.4.6.

$$
\begin{aligned}
\sqrt{n}(\hat{\alpha}_{ii}-\alpha_{ii}) &= \sqrt{n}\sum_{u=1}^{m} L_{u,i}(\log \bar{Z}_{n,i}^{u}-\alpha_{ii}\log u) \\
=&\sqrt{n}\sum_{u=1}^{m} L_{u,i}\log(1+T_{n,i}^{u})+\sqrt{n}\sum_{u=1}^{m} L_{u,i}(\log \mathbb{E}\bar{Z}_{n,i}^{u}-\alpha_{ii}\log u) \\
=&\sqrt{n}\sum_{u=1}^{m} L_{u,i}T_{n,i}^{u}(1+e_{n,i}^{u})+O(n^{-\beta_{ii}+\frac{1}{2}}) \\
=&\sqrt{n}L_i^{\top}T_{n,i}+\sqrt{n}\sum_{u=1}^{m} L_{u,i}T_{n,i}^{u}e_{n,i}^{u}+O(n^{-\beta_{ii}+\frac{1}{2}}),
\end{aligned}
\tag{3.6.54}
$$

where $e_{n,i}^{u}\to 0$ if $T_{n,i}^{u}\to 0$.

By Theorem 3.4.3, we have

$$
\sqrt{n}\begin{pmatrix} L_1^{\top}T_{n,1} \\ L_2^{\top}T_{n,2} \end{pmatrix} \xrightarrow{d} N(\mathbf{0},\Sigma_{\alpha}),
\tag{3.6.55}
$$

where

$$
\Sigma_{\alpha}=\begin{pmatrix} \tilde{L}_1^{\top}\Phi_{0,11}\tilde{L}_1 & \tilde{L}_1^{\top}\Phi_{0,12}\tilde{L}_2 \\ \tilde{L}_2^{\top}\Phi_{0,21}\tilde{L}_1 & \tilde{L}_2^{\top}\Phi_{0,22}\tilde{L}_2 \end{pmatrix}.
\tag{3.6.56}
$$

Especially, if $\frac{\alpha_{11}+\alpha_{22}}{2}<\alpha_{12}$, $\Phi_{0,12}=\Phi_{0,21}=\mathbf{0}$ and hence $L_1^{\top}T_{n,1}$ and $L_2^{\top}T_{n,2}$ are asymptotically independent.

Since $\sqrt{n}L_{u,i}T_{n,i}^{u}$ convergence to normal distribution and $e_{n,i}\to 0$ as $n\to\infty$, we have $\sqrt{n}L_{u,i}T_{n,i}^{u}e_{n,i}=o_p(1)$ and hence $\sqrt{n}\sum_{u=1}^{m}L_{u,i}T_{n,i}^{u}e_{n,i}^{u}=o_p(1)$. Therefore,

$$
\sqrt{n}(\hat{\alpha}_{ii}-\alpha_{ii})=\sqrt{n}L_i^{\top}T_{n,i}+o_p(1)+O(n^{-\beta_{ii}+\frac{1}{2}}).
\tag{3.6.57}
$$

$\square$

Lemma 3.6.1. *Show that for any fixed* $k \in \mathbb{Z}^+$,

$$\sum_{u=1}^{m} L_{u,i}\mathbb{E}[\log^k(1+T_{n,i}^u), |T_{n,i}^u| > \xi] \leq cn^{1/2}e^{-c\sqrt{n}}, \tag{3.6.58}$$

where c *is a constant.*

Proof of Lemma 3.6.1. By Holder's inequality, we have

$$\mathbb{E}[\log^k(1+T_{n,i}^u), |T_{n,i}^u| > \xi] \leq \mathbb{E}^{1/2}[\log^{2k}(1+T_{n,i}^u)]\mathbb{P}^{1/2}(|T_{n,i}^u| > \xi). \tag{3.6.59}$$

First, we find a upper bound for $\mathbb{P}(|T_{n,i}^u| > \xi)$.

Let $U = (U_1, ..., U_n)$ where $U_i \overset{iid}{\sim} N(0,1)$. Let $Y_n = (Y_{n,i}^u(1), ..., Y_{n,i}^u(n))$. Denote by $Var(Y_n) := \Sigma_n = (\sigma_{n,ii}^{uu}(j-k))_{j,k=1}^n$. Let $\Lambda_n = diag(\lambda_i)$ where $\lambda_i, i = 1, ..., n$ are the eigenvalues of Σ_n. Then we have

$$\bar{Z}_{n,i}^u = \frac{1}{n}Y_n^\top Y_n \overset{d}{=} \frac{1}{n}U^\top \Lambda_n U. \tag{3.6.60}$$

Denote by $||\Lambda_n||_2$ and $||\Lambda_n||_F$ the l_2 norm and Frobenius norm of Λ_n respectively. Indeed,

$$||\Lambda_n||_2 = \max_{1\leq j\leq n} \lambda_j, \; ||\Lambda_n||_F = \sqrt{\sum_{j=1}^{n} \lambda_j^2}. \tag{3.6.61}$$

It is easy to see $n\mathbb{E}\bar{Z}_{n,i}^u = \mathbb{E}U^\top \Lambda_n U = tr(\Lambda_n)$ and hence $tr(\Lambda_n)/n \to C_3 u^{\alpha_{ii}}$ where C_3 is a constant. Further, Since

$$||\Lambda_n||_F^2 = tr(\Lambda_n^2) = tr(\Sigma_n^2) = \sum_{j=1,k=1}^{n} (\sigma_{n,ii}^{uu}(j-k))^2, \tag{3.6.62}$$

and

$$\phi_{n,ii}^{uu} = Var(\bar{Z}_{n,ii}^{u}) = \frac{2}{n^2} \sum_{j=1,k=1}^{n} (\sigma_{n,ii}^{uu}(j-k))^2, \tag{3.6.63}$$

we have

$$||\Lambda_n||_F^2 = \frac{n^2}{2}\phi_{n,ii}^{uu} \asymp \frac{n}{2}\phi_{0,ii}^{uu}. \tag{3.6.64}$$

Applying Hanson and Wright's inequality [HW71], we have

$$\begin{aligned} &\mathbb{P}(|T_{n,i}^{u}| > \xi) = \mathbb{P}(|U^\top \Lambda_n U - tr(\Lambda_n)| > tr(\Lambda_n)\xi) \\ \leq &\exp\left\{ -\min\left(C_1\xi\frac{tr(\Lambda_n)}{||\Lambda_n||_2}, C_2\xi^2\frac{(tr(\Lambda_n))^2}{||\Lambda_n||_F^2} \right) \right\} \end{aligned} \tag{3.6.65}$$

where C_1, C_2 are positive constants independent of Λ_n, n and ξ.

Since $||\Lambda_n||_2 \leq ||\Lambda_n||_F$ and $||\Lambda_n||_F^2/n \to \frac{1}{2}\phi_{0,ii}^{u}$, we have

$$\frac{tr(\Lambda_n)}{||\Lambda_n||_2} = \sqrt{n}\frac{tr(\Lambda_n)/n}{||\Lambda_n||_2/\sqrt{n}} \gtrsim \sqrt{2}C_3u^{\alpha_{ii}}(\phi_{0,ii}^{uu})^{-1/2}\sqrt{n}, \text{ as } n \to \infty, \tag{3.6.66}$$

and

$$\frac{(tr(\Lambda_n))^2}{||\Lambda_n||_F^2} \asymp 2C_3^2u^{2\alpha_{ii}}(\phi_{0,ii}^{uu})^{-1}n. \tag{3.6.67}$$

Hence, when $n \to \infty$, (3.6.65) decays exponentially with rate $\sqrt{n}$, specifically,

$$\mathbb{P}(|T^u_{n,i}| > \xi) \leq e^{-\sqrt{2}C_1 C_3 u^{\alpha_{ii}} (\phi^{uu}_{0,ii})^{-1/2} \xi \sqrt{n}}, \text{ as } n \to \infty. \tag{3.6.68}$$

Second, we prove the upper bound of $\mathbb{E}\log^{2k}(1+T_{n,i})$. It is easy to see that

$$\mathbb{E}\log^{2k}(1+T_{n,i}) \leq 2^{2k-1}\big(\mathbb{E}\log^{2k}\bar{Z}^u_{n,i} + \log^{2k}\mathbb{E}\bar{Z}^u_{n,i}\big). \tag{3.6.69}$$

For any fixed $k \in \mathbb{Z}^+$, there exists c_k such that $\log^{2k} x \leq x^2, \forall x > c_k$. Since $\mathbb{E}\bar{Z}^u_{n,i} \to C_3 u^{\alpha_{ii}}$,

$$\mathbb{E}(\log^{2k}\bar{Z}^u_{n,i}; \bar{Z}^u_{n,i} > c_k) \leq \mathbb{E}(\bar{Z}^u_{n,i})^2 = (\mathbb{E}\bar{Z}^u_{n,i})^2 + Var(\bar{Z}^u_{n,i}) \to C_3^2 u^{2\alpha_{ii}}. \tag{3.6.70}$$

So $\mathbb{E}(\log^{2k}\bar{Z}^u_{n,i}; \bar{Z}^u_{n,i} > c_k)$ is uniformly bounded and we only needs to check $\mathbb{E}(\log^{2k}\bar{Z}^u_{n,i}; \bar{Z}^u_{n,i} \leq c_k)$.

Let $U^2_{\min} = \min_{1\leq i\leq n} U_i^2$. For n large,

$$\bar{Z}^u_{n,i} = \frac{1}{n}\sum_{i=1}^n \lambda_i U_i^2 \geq \frac{tr(\Lambda_n)}{n} U^2_{\min} \geq \frac{1}{2} C_3 u^{\alpha_{ii}} U^2_{\min} \triangleq C U^2_{\min}. \tag{3.6.71}$$

Let $f_n(x)$ be the density function of $U^2_{\min}$, that is

$$f_n(x) := \mathbb{P}(U^2_{\min} \in dx) = \frac{2n}{\sqrt{2\pi}} e^{-x/2}\Big(2\int_{\sqrt{x}}^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy\Big)^{n-1} dx. \tag{3.6.72}$$

It is easy to check that $f_n(x) \leq \frac{2n}{\sqrt{2\pi}}$. Hence, when n is large, we obtain

$$\mathbb{E}(\log^{2k}\bar{Z}^u_{n,i}; \bar{Z}^u_{n,i} \leq c_k) \leq \mathbb{E}\big(\log^{2k}\left(CU^2_{\min}\right); U^2_{\min} \leq 1/C\big)$$

$$\int_0^{c_k} \log^{2k}(Cx) f_n(x)dx \leq \frac{2n}{\sqrt{2\pi}C} \int_0^1 \log^{2k} y dy \leq cn. \tag{3.6.73}$$

By (3.6.69), (3.6.70) and (3.6.73), we obtain $\mathbb{E}\log^{2k}(1 + T_{n,i}) \leq cn$ and hence when n is large,

$$\mathbb{E}[\log^k(1 + T_{n,i}^u), |T_{n,i}^u| > \xi] \leq cn^{1/2}e^{-c\sqrt{n}}. \tag{3.6.74}$$

$\square$

Chapter 4

A Bayesian functional data model for coupling high-dimensional LiDAR and forest variables over large geographic domains

Recent advances in remote sensing, specifically Light Detection and Ranging (LiDAR) sensors, provide the data needed to quantify forest variables at a fine spatial resolution over large domains. We define a framework to couple high-dimensional and spatially indexed LiDAR signals with forest variables using a fully Bayesian functional spatial data analysis. The proposed modeling framework is illustrated by a simulated study and by analyzing LiDAR and spatially coinciding forest inventory data collected on the Penobscot Experimental Forest, Maine.

4.1 Introduction

Linking long-term forest inventory with air- and space-borne Light Detection and Ranging (LiDAR) datasets via regression models offers an attractive approach to mapping forest above-ground biomass (AGB) at stand, regional, continental, and global scales. LiDAR data have shown great potential for use in estimating spatially explicit forest variables, including AGB, over a range of geographic scales [AHV+09, BMF+13, FBM11, IDUA13, MMT11, Næs11, NNR+13]. Encouraging results from these and many other studies have spurred massive investment in new LiDAR sensors, sensor platforms, as well as extensive campaigns to collect field-based calibration data. For example, ICESat-2—planned for launch in 2017—will be equipped with a LiDAR sensor able to gather data from space at unprecedented spatial resolutions [AZB+10]. As currently proposed, ICESat-2 will be a photon-counting sensor capable of recording measurements on a ≈70 cm footprint [ICE15]. The Global Ecosystem Dynamics Investigation LiDAR (GEDI) will be an International Space Station mounted system capable of producing 25 m diameter footprint waveforms and is scheduled to be operational in 2018 [GED14]. One of GEDI's core objectives is to quantify the distribution of AGB at a fine spatial resolution. NASA Goddard's LiDAR, Hyper-spectral, and Thermal (G-LiHT) imager is an air-borne platform developed, in part, to examine how future space-originating LiDAR, e.g., ICESat-2, GEDI, or other platforms, may be combined with field-based validation measurements to build predictive models for AGB and other forest variables [AAWN13, CCN+13].

In order to effectively extract information from these high-dimensional massive datasets, we need a modeling framework to capture within and among LiDAR signal/forest variable association within and across locations. However, the computational complexity of such models increases in cubic order with the number of spatial locations and the dimension of the LiDAR signal, and the number of forest variables—a characteristic common to multivariate spatial process models. In this chapter, we propose a modeling framework that explicitly: 1) reduces the dimensionality of signals in an optimal way (i.e., preserves the information that describes the maximum variability in response variable); 2) propagates uncertainty in data and parameters through to prediction, and; 3) acknowledges and leverages spatial dependence among the derived regressors and model residuals to meet statistical assumptions and improve prediction.

The rest of this chapter is organized as follows. Section 4.2 gives a review on Gaussian predictive process and modified Gaussian predictive process. The join model that coupling the LiDAR signal and the forest variables is proposed in Section 4.3. In Section 4.4, we complete the hierarchical specifications and outlines the Gibbs sampler for fitting the model. Spatial predictions for forest variables are derived in Section 4.5. Section 4.6 illustrates with simulation and forestry data analysis.

4.2 Preliminary: Modified Gaussian predictive process

A Gaussian random field $\{w(s), s \in D \subset \mathbb{R}^N\}$ is usually used to model the residual of spatial point referenced data. Let $C_w(s,t;\theta) := \mathrm{Cov}(\mathrm{w(s)},\mathrm{w(t)})$ be the covariance function of $w(\cdot)$ with parameters θ. Assume that the data are observed in n locations, say, $s_1, ..., s_n$. Estimating the parameters always needs inverting a $n \times n$ covariance matrix, which involves $O(n^3)$ flops. When the sample size n is very big, it is computationally very expensive and even infeasible.

To address this issue, [BGFS08] introduced the Gaussian predictive process model, which is a degenerate Gaussian random field obtained by projecting the parent random field to a lower-dimensional subspace. Specifically, by choosing a set of "knots" $\mathcal{S} = \{s_1^*, ..., s_r^*\} \subset D$, they defined the Gaussian predictive process

$$\tilde{w}(s) = \mathbb{E}(w(s)|w(s_1^*), ..., w(s_r^*)). \tag{4.2.1}$$

Let $c(s;\theta) = [C_w(s, s_j^*;\theta)]_{j=1}^r$ and $C^*(\theta) = [C_w(s_i^*, s_j^*)]_{i,j=1}^r$. The covariance function of $\tilde{w}(s)$ can be derived directly from its parent process, that is

$$\mathrm{Cov}(\tilde{w}(s), \tilde{w}(t)) = c^\top(s;\theta)C^{*-1}(\theta)c(t;\theta). \tag{4.2.2}$$

Inverting the corresponding covariance matrix only requires $O(nr^2)$ flops with $r << n$.

[FSBG09] pointed out that the predictive process underestimates the variance of the parent random field $w(s)$ since $\forall s \in D$,

$$\begin{aligned} &\mathrm{Var(w(s)) - Var(\tilde{w}(s)) = Var(w(s)|w(s_1^*), ..., w(s_r^*))} \\ &= C_w(s,s;\theta) - c^\top(s;\theta)C^{*-1}(\theta)c(s;\theta) \geq 0. \end{aligned} \tag{4.2.3}$$

As a consequence, the nugget variance in spatial regression model is usually overestimated by absorbing the variability dropped by the predictive process. To remedy this problem, [FSBG09] proposed the **modified Gaussian predictive process** by adding a Gaussian noise to the predictive process. Specifically, they defined

$$\ddot{w}(s) = \tilde{w}(s) + \tilde{\epsilon}(s), \tag{4.2.4}$$

where $\tilde{\epsilon}(s) \overset{ind}{\sim} N(0, C_w(s,s;\theta) - c^\top(s;\theta)C^{*-1}(\theta)c(s;\theta))$ is a spatially independent Gaussian random field with varying marginal variance. Hence, the modified Gaussian predictive process has the same marginal variance as the parent process.

4.3 The model

Let $D \subset \mathbb{R}^2$ be a spatial domain and let s be a generic point in D. At location s, the outcome variable $y(s)$ denotes the above-ground biomass. Let $x \in [0, M] \subset \mathbb{R}^+$ be the height from the ground with M the maximum height. At location (s, x), the outcome variable $z(s, x)$ denotes the strength of LiDAR signal.

Assume we've observed both y and z at a set of locations $\mathcal{S} = \{s_1, ..., s_n\}$. For each location s_i, z is measured at height $x_1, x_2, ..., x_{n_x}$. Moreover, we observed the LiDAR signal z in many other locations where y have not been measured.

4.3.1 Modified Gaussian predictive model for z

The signal $z(s, x)$ can be modeled as follows,

$$z(s, x) = \mu_z(s, x; \beta) + u(s, x) + \epsilon_z(s, x), \tag{4.3.1}$$

where μ_z is the mean function, $u(s, x)$ is the random effect which is a Gaussian random field on $\mathbb{R}^3$ and $\epsilon_z(s, x)$ is the nugget effect.

Assume that the nuggets effect $\epsilon_z(s, x) \overset{ind}{\sim} N(0, \tau_z^2(x))$, which means the variance of the nugget is independent across locations.

Denote by $C_u(s, t, x, y; \theta_u) := \mathrm{Cov}[u(s, x), u(t, y)]$ the covariance function of random effect u. We approximate the parent model by modified Gaussian predictive processes within locations. Assume that $\{x_1^*, ..., x_{n_x^*}^*\}$ are the height knots at every location. Let

$$u^*(s) = (u(s, x_1^*), ..., u(s, x_{n_x^*}^*))^\top,$$

whose covariance at location s is given by

$$V_{u^*(s)}(\theta_u) := \begin{pmatrix} C_u(s, s, x_1^*, x_1^*; \theta_u) & \cdots & C_u(s, s, x_1^*, x_{n_x^*}^*; \theta_u) \\ C_u(s, s, x_2^*, x_1^*; \theta_u) & \cdots & C_u(s, s, x_2^*, x_{n_x^*}^*; \theta_u) \\ \vdots & \ddots & \vdots \\ C_u(s, s, x_{n_x^*}^*, x_1^*; \theta_u) & \cdots & C_u(s, s, x_{n_x^*}^*, x_{n_x^*}^*; \theta_u) \end{pmatrix}, \tag{4.3.2}$$

and let

$$f(s, x; \theta_u) := (C_u(s, s, x, x_1^*; \theta_u), ..., C_u(s, s, x, x_{n_x^*}^*; \theta_u))^\top. \tag{4.3.3}$$

So the **modified Gaussian predictive process model for** z within locations is

$$z(s, x) = \mu_z(s, x; \beta) + f^\top(s, x; \theta_u) V_{u^*(s)}^{-1}(\theta_u) u^*(s) + \epsilon_z^*(s, x), \tag{4.3.4}$$

where

$$\epsilon_z^*(s; x) \stackrel{ind}{\sim} N(0, \tilde{\sigma}_u^2(s, x; \theta_u) + \tau_z^2(x)), \tag{4.3.5}$$

with

$$\tilde{\sigma}_u^2(s, x; \theta_u) = C_u(s, s, x, x; \theta_u) - f^\top(s, x; \theta_u) V_{u^*(s)}^{-1}(\theta_u) f(s, x; \theta_u). \tag{4.3.6}$$

4.3.2 Joint model of y and Z

Denote by

$$Z(s) = \begin{pmatrix} z(s, x_1) \\ z(s, x_2) \\ \vdots \\ z(s, x_{n_x}) \end{pmatrix}, \epsilon_Z^*(s) = \begin{pmatrix} \epsilon_z^*(s; x_1) \\ \epsilon_z^*(s; x_2) \\ \vdots \\ \epsilon_z^*(s; x_{n_x}) \end{pmatrix}, F(s; \theta_u) = \begin{pmatrix} f^\top(s, x_1; \theta_u) \\ f^\top(s, x_2; \theta_u) \\ \vdots \\ f^\top(s, x_{n_x}; \theta_u) \end{pmatrix},$$

and

$$\mu_Z(s; \beta) = \begin{pmatrix} \mu_z(s, x_1; \beta) \\ \mu_z(s, x_2; \beta) \\ \vdots \\ \mu_z(s, x_{n_x}; \beta) \end{pmatrix}.$$

At location s, we couple the LiDAR signal Z and the above-ground biomass y through the modified predictive processes $u^*(s)$. The **joint model for y and Z** is given below,

$$\begin{pmatrix} Z(s) \\ y(s) \end{pmatrix} = \begin{pmatrix} \mu_Z(s; \beta) \\ \mu_y(s; \eta) \end{pmatrix} + \begin{pmatrix} F(s; \theta_u) V_{u^*(s)}^{-1}(\theta_u) & \mathbf{0} \\ \alpha_u^\top & 1 \end{pmatrix} \begin{pmatrix} u^*(s) \\ v(s) \end{pmatrix} + \begin{pmatrix} \epsilon_Z^*(s) \\ \epsilon_y(s) \end{pmatrix}, \tag{4.3.7}$$

where $\alpha_u \in \mathbb{R}^{n_x^*}$, μ_y is the mean function of y, $v(s)$ is the random effect and ϵ_y is the nugget effect. Assume that v, ϵ_y are independent of u and ϵ_z. $v(\cdot)$ is a Gaussian random field on $\mathbb{R}^2$ and $\epsilon_y(s) \overset{ind}{\sim} N(0, \tau_y^2)$.

When the number of locations n is big, model estimation and prediction is computationally inefficient. So we use the modified Gaussian predictive process to approximate the joint model for y and Z by choosing the spatial knots. Let $\{s_1^*, s_2^*, ..., s_{n_z^*}^*\}$ and $\{t_1^*, t_2^*, ..., t_{n_y^*}^*\}$ be the spatial knots for signal Z and y respectively.

First, we approximate the spatial random effect $u^*(s)$. Let $u^* = (u^{*T}(s_1^*), ..., u^{*T}(s_{n_z^*}^*))^\top$ and $G_{u^*}(s;\theta_u)^\top$ be the $n_x^* \times n_x^* n_z^*$ block matrix with $(1,j)$-th block being the cross covariance $C_{u^*}(s, s_j^*; \theta_u) := \text{Cov}[u^*(s), u^*(s_j^*)]$, $j = 1, 2, .., n_z^*$, that is

$$G_{u^*}(s;\theta_u)^\top = (C_{u^*}(s, s_1^*; \theta_u), C_{u^*}(s, s_2^*; \theta_u), ..., C_{u^*}(s, s_{n_z^*}^*; \theta_u)). \tag{4.3.8}$$

Denote by $V_{u^*}(\theta_u)$ the $n_x^* n_z^* \times n_x^* n_z^*$ block matrix whose (i,j)-th block is $C_{u^*}(s_i^*, s_j^*; \theta_u)$. We approximate $u^*(s)$ by the modified Gaussian predictive process below,

$$\tilde{u}^*(s) = G_{u^*}(s;\theta_u)^\top V_{u^*}^{-1}(\theta_u) u^* + \epsilon_{u^*}(s), \tag{4.3.9}$$

where

$$\epsilon_{u^*}(s) \sim N(\mathbf{0}, \Sigma_{u^*}(s;\theta_u)), \tag{4.3.10}$$

with

$$\Sigma_{u^*}(s;\theta_u) = C_{u^*}(s, s; \theta_u) - G_{u^*}(s;\theta_u)^\top V_{u^*}^{-1} G_{u^*}(s;\theta_u). \tag{4.3.11}$$

Next, we approximate $v(s)$. Let $g_v(s;\theta_v)^\top$ be the $1 \times n_y^*$ vector whose j-th element is $C_v(s, s_j^*; \theta_v) := \text{Cov}[v(s), v(s_j^*)]$ and denote by $V_{v^*}(\theta_v)$ the $n_y^* \times n_y^*$ matrix whose (i,j)-th element is $C_v(s_i^*, s_j^*; \theta_v)$. $v(s)$ is approximated by

$$\tilde{v}(s) = g_v(s;\theta_v)^\top V_{v^*}^{-1}(\theta_v) v^* + \epsilon_v(s), \tag{4.3.12}$$

where $\epsilon_v(s) \sim N(0, \sigma_v^2(s;\theta_v))$ with

$$\sigma_v^2(s;\theta_v) = C_v(s,s;\theta_v) - g_v(s;\theta_v)^\top V_{v^*}^{-1}(\theta_v) g_v(s;\theta_v). \tag{4.3.13}$$

Therefore, $Z(s)$ in (4.3.7) can be approximated by

$$\begin{aligned} Z(s) &\approx \mu_Z(s;\beta) + F(s;\theta_u) V_{u^*(s)}^{-1}(\theta_u) \tilde{u}^*(s) + \epsilon_Z^*(s) \\ &\triangleq \mu_Z(s;\beta) + H(s;\theta_u) u^* + e_Z(s), \end{aligned} \tag{4.3.14}$$

where

$$\begin{aligned} &H(s;\theta_u) := F(s;\theta_u) V_{u^*(s)}^{-1}(\theta_u) G_{u^*}(\theta_u; s)^\top V_{u^*}^{-1}(\theta_u), \\ &e_Z(s) = F(s;\theta_u) V_{u^*(s)}^{-1}(\theta_u) \epsilon_{u^*}(s) + \epsilon_Z^*(s) \sim N(\mathbf{0}, D_{e_Z}(s)), \end{aligned}$$

with

$$D_{e_Z}(s) = F(s;\theta_u) V_{u^*(s)}^{-1}(\theta_u) \Sigma_{u^*}(s;\theta_u) V_{u^*(s)}^{-1}(\theta_u) F(s;\theta_u)^\top + \oplus_{j=1}^{nx} (\tilde{\sigma}_u^2(s, x_j;\theta_u) + \tau_z^2(x)).$$

Also, $y(s)$ defined in (4.3.7) can be approximated by

$$y(s) \approx \mu_y(s;\eta) + \alpha_u^\top \tilde{u}^*(s) + \tilde{v}(s) + \epsilon_y(s)$$

$$\triangleq \mu_y(s,\eta) + I(s;\theta_u)u^* + J(s;\theta_v)v^* + e_y(s), \tag{4.3.15}$$

where

$$\begin{aligned} & I(s;\theta_u) = \alpha_u^\top G_{u^*}(s;\theta_u)^\top V_{u^*}^{-1}(\theta_u), J(s;\theta_v) = g_v(s;\theta_v)^\top V_{v^*}^{-1}(\theta_v), \\ & e_y(s) := \alpha_u^\top \epsilon_{u^*}(s) + \epsilon_v(s) + \epsilon_y(s) \sim N(\mathbf{0}, \sigma_{e_y}^2(s)), \end{aligned} \tag{4.3.16}$$

with

$$\sigma_{e_y}^2(s) = \alpha_u^\top \Sigma_{u^*}(s;\theta_u)\alpha_u + \sigma_v^2(s;\theta_v) + \tau_y^2. \tag{4.3.17}$$

So our **final model** for Z and y can be rewritten as

$$\begin{pmatrix} Z(s) \\ y(s) \end{pmatrix} = \begin{pmatrix} \mu_Z(s;\beta) \\ \mu_y(s;\eta) \end{pmatrix} + \begin{pmatrix} H(s;\theta_u) & \mathbf{0} \\ I(s;\theta_u) & J(s;\theta_v) \end{pmatrix} \begin{pmatrix} u^* \\ v^* \end{pmatrix} + \begin{pmatrix} e_Z(s) \\ e_y(s) \end{pmatrix}. \tag{4.3.18}$$

Let $X_Z(s) \in \mathbb{R}^{nx \times p}$ and $X_y(s) \in \mathbb{R}^q$ be the predictors for the signal Z and y respectively. Assume that $\mu_Z(s;\beta) = X_Z^\top(s)\beta$ and $\mu_y(s;\eta) = X_y(s)^\top \eta$. Then,

$$\begin{pmatrix} Z(s) \\ y(s) \end{pmatrix} = \begin{pmatrix} X_Z(s)^\top\beta \\ X_y(s)^\top\eta \end{pmatrix} + \begin{pmatrix} H(s;\theta_u) & \mathbf{0} \\ I(s;\theta_u) & J(s;\theta_v) \end{pmatrix} \begin{pmatrix} u^* \\ v^* \end{pmatrix} + \begin{pmatrix} e_Z(s) \\ e_y(s) \end{pmatrix}. \tag{4.3.19}$$

4.3.3 Specification of the random effect of Z and y

- ***Covariance function for*** $u(s,x)$***:*** [Gne02] introduced a class of nonseparable stationary covariance function for space-time model on $\mathbb{R}^d \times \mathbb{R}$. Specifically, when $d = 2$,

$$\begin{aligned} C(s_1, s_2; x_1, x_2) =& \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)(a|x_1 - x_2|^{2\kappa} + 1)^{\delta+\gamma}} \left(\frac{c \parallel s_1 - s_2 \parallel}{(a|x_1 - x_2|^{2\kappa} + 1)^{\gamma/2}} \right)^{\nu} \\ & \times K_\nu \left(\frac{c \parallel s_1 - s_2 \parallel}{(a|x_1 - x_2|^{2\kappa} + 1)^{\gamma/2}} \right) \end{aligned} \tag{4.3.20}$$

Here, we use a simplified version to model the covariance of $u(s,x)$ by fixing $\kappa = 1, \nu = \frac{1}{2}, \delta = 0$, i.e.,

$$C_u(s_1, s_2, x_1, x_2; \theta_u) := \frac{\sigma_u^2}{(a|x_1 - x_2|^2 + 1)^{\gamma}} \exp\left(- \frac{c \parallel s_1 - s_2 \parallel}{(a|x_1 - x_2|^2 + 1)^{\gamma/2}} \right) \tag{4.3.21}$$

where $\theta_u = (\sigma_u^2, a, \gamma, c), \sigma_u^2, a, c > 0$ and $\gamma \in [0, 1]$.

- ***Covariance function for*** $v(s)$***:*** We employ the exponential covariance function for $v(s)$, i.e.,

$$C_v(s_1, s_2; \theta_v) = \sigma_v^2 \exp\{-\phi_v \parallel s_1 - s_2 \parallel\}, \tag{4.3.22}$$

where $\theta_v = (\sigma_v^2, \phi_v)$.

4.4 Bayesian implementation and computational issue

4.4.1 Data equation

For fitting the final model in (4.3.19), we form the data equation in this section. Denote by $\theta = (\theta_u, \theta_v)$. Let

$$O(s) = \begin{pmatrix} Z(s) \\ y(s) \end{pmatrix}, B(s;\theta) = \begin{pmatrix} H(s;\theta_u) & \mathbf{0} \\ I(s;\theta_u) & J(s;\theta_v) \end{pmatrix}, b = \begin{pmatrix} \beta \\ \eta \end{pmatrix}$$

$$X(s) = \begin{pmatrix} X_Z(s)^\top & 0 \\ 0 & X_y(s)^\top \end{pmatrix}, w^* = \begin{pmatrix} u^* \\ v^* \end{pmatrix}, e(s) = \begin{pmatrix} e_Z(s) \\ e_y(s) \end{pmatrix} \tag{4.4.1}$$

The data model for (4.3.19) can be written as

$$O(s_i) = X(s_i)b + B(s_i;\theta)w^* + e(s_i), i = 1, 2, ..., n. \tag{4.4.2}$$

The matrix form of the above model is

$$O = Xb + B(\theta)w^* + e, \tag{4.4.3}$$

where

$$O = \begin{pmatrix} O(s_1) \\ \vdots \\ O(s_n) \end{pmatrix}, B(\theta) = \begin{pmatrix} B(s_1;\theta) \\ \vdots \\ B(s_n;\theta) \end{pmatrix}, e = \begin{pmatrix} e(s_1) \\ \vdots \\ e(s_n) \end{pmatrix}, X = \begin{pmatrix} X(s_1) \\ \vdots \\ X(s_n) \end{pmatrix}. \tag{4.4.4}$$

Here,

$$e \sim N(\mathbf{0}, D_e), \tag{4.4.5}$$

where

$$D_e = \begin{pmatrix} D_0(s_1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & D_0(s_2) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & D_0(s_n) \end{pmatrix}, \tag{4.4.6}$$

with

$$D_0(s_i) = \begin{pmatrix} D_{e_Z}(s_i) & \mathbf{0} \\ \mathbf{0} & \sigma^2_{ey}(s_i) \end{pmatrix}, i = 1, 2, ..., n. \tag{4.4.7}$$

And

$$w^* \sim MVN(\mathbf{0}, \Sigma_{w^*}), \tag{4.4.8}$$

where

$$\Sigma_{w^*} = \begin{pmatrix} V_{u^*}(\theta_z) & \mathbf{0} \\ \mathbf{0} & V_{v^*}(\theta_v) \end{pmatrix} \tag{4.4.9}$$

4.4.2 Prior specification and full conditional sampling

Denote by Ω all the parameters in the model, which is

$$\Omega = \{\beta, \eta, \alpha_u, \theta, \tau_y^2, \tau_z^2(x_i), i = 1, ..., n_x\}. \tag{4.4.10}$$

First, let us complete hierarchical specifications. The priors for all the parameters are given as follows.

$$\begin{aligned}
&\beta \sim N(\mathbf{0}, \mathrm{diag}(10^4)),\ \eta \sim \mathrm{N}(\mathbf{0}, \mathrm{diag}(10^4)),\ \alpha_{\mathrm{u,i}} \sim \mathrm{N}(0, \mathrm{diag}(10^4)), \mathrm{i} = 1, ..., \mathrm{n}_{\mathrm{x}}^*,\\
&\sigma_u^2 \sim IG(2, b_{\sigma_u}), \gamma \sim U(0,1),\ a \sim U(0, a_{max}),\ c \sim U(0, c_{max}),\\
&\sigma_v^2 \sim IG(2, b_{\sigma_v}),\ \phi_v \sim U(-\log(0.05)/d_{s,max}, -\log(0.01)/d_{s,min}),\\
&\tau_z^2(x_i) \sim IG(2, b_{\tau_z}), i = 1, 2, ..., n_x,\ \tau_y^2 \sim IG(2, b_{\tau_y}).
\end{aligned} \tag{4.4.11}$$

We assign large enough numbers for a_{max} and c_{max}. The specification of hyperparameters for ϕ_v follow from [RB13], where $d_{s,min}, d_{s,max}$ are minimum and maximum distance across all the locations. $b_{\sigma_u}, b_{\sigma_v}, b_{\tau_z}$, and b_{τ_y} are assigned from the empirical semivariogram (see, e.g., [BGFS08]).

Denote by $\Theta := \Omega \setminus b = (\alpha_u, \theta, \tau_z^2(x_i), \tau_y^2)$. We can integrate out the random effect w^* and obtain the marginalized likelihood,

$$[O \mid \Theta, b] \sim MVN(Xb, B(\theta)\Sigma_{w^*}B^\top(\theta) + D_e). \tag{4.4.12}$$

So the posterior distribution of the model is

$$[\Theta, b \,|\, O] \propto MVN(Xb, B(\theta)\Sigma_{w^*}B^\top(\theta) + D_e) \times [\Theta] \times [b]. \tag{4.4.13}$$

Model fitting employs a Gibbs sampler with Metropolis steps.

1. **Update** $b = (\beta, \eta)^\top$. Let

$$\mu_b = (\mu_\beta, \mu_\eta)^\top,\ \Sigma_{O|\Omega} = B(\theta)\Sigma_{w^*}B^\top(\theta) + D_e,\ \text{and } \Sigma_b = \begin{pmatrix} \Sigma_\beta & 0 \\ 0 & \Sigma_\eta \end{pmatrix} \tag{4.4.14}$$

   The full conditional density for b is

$$[b|O, \Theta] \sim N(\mu_{|\cdot}, \Sigma_{b|\cdot}), \tag{4.4.15}$$

   where

$$\Sigma_{b|\cdot} = \left(\Sigma_b^{-1} + X^\top \Sigma_{O|\Omega}^{-1} X\right)^{-1} \text{ and } \mu_{b|\cdot} = \Sigma_{b|\cdot}\left(\Sigma_b^{-1}\mu_b + X^\top \Sigma_{O|\Omega}^{-1} O\right) \tag{4.4.16}$$

2. **Update Θ.** We employ the block random walk Metropolis method to sample Θ.

   - **Support on $\mathbb{R}$.** We'll do some log-type transformation to make sure the support of all the parameters is $(-\infty, \infty)$. Specifically, for the parameters with Inverse Gamma prior, we just do the log-transformation, i.e.,

$$\tilde{\sigma}_u^2 = \log \sigma_u^2, \tilde{\sigma}_v^2 = \log \sigma_v^2, \tilde{\tau}_z^2(x_i) = \log \tau_z^2(x_i),\ \text{and } \tilde{\tau}_y^2 = \log \tau_y^2. \tag{4.4.17}$$

For the parameters with uniform prior, we do the logit-transformation. Specifically,

$$\begin{aligned}\tilde{\phi}_v &= \log\left(\frac{\phi_v - \phi_{v,min}}{\phi_{v,max} - \phi_v}\right), \tilde{a} = \log\left(\frac{a}{a_{max} - a}\right),\\ \tilde{\gamma} &= \log\left(\frac{\gamma}{1-\gamma}\right), \tilde{c} = \log\left(\frac{c}{c_{max} - c}\right),\end{aligned} \tag{4.4.18}$$

where $\phi_{v,min} = -\log(0.05)/d_{s,max},\ \phi_{v,max} = -\log(0.01)/d_{s,min}$.

- **Log-likelihood of $\tilde{\Theta}$.** Denote by $\tilde{\Theta}$ the parameters of Θ after transformation. Given O, b, the log-likelihood of $\tilde{\Theta}$ is proportional to

$$\begin{aligned}&\log p(\tilde{\Theta}|O,b)\\ \propto &-\frac{1}{2}\log|\Sigma_{O|\Omega}| - \frac{1}{2}(O - Xb)^\top \Sigma_{O|\Omega}^{-1}(O - Xb) - \sum_{i=1}^{n_x^*} \frac{\alpha_{u,i}^2}{2\sigma_{\alpha_i}^2}\\ &- a_{\sigma_u}\tilde{\sigma}_u^2 - a_{\sigma_v}\tilde{\sigma}_v^2 - a_{\tau_z}\sum_{i=1}^{n_x}\tilde{\tau}_z^2(x_i) - a_{\tau_y}\tilde{\tau}_y^2\\ &- b_{\sigma_u}e^{-\tilde{\sigma}_u^2} - b_{\sigma_v}e^{-\tilde{\sigma}_v^2} - b_{\tau_z}\sum_{i=1}^{n_x} e^{-\tilde{\tau}_z^2(x_i)} - b_{\tau_y}e^{-\tilde{\tau}_y^2}\\ &+ \tilde{a} + \tilde{\gamma} + \tilde{c} + \tilde{\phi}_v - 2\log\left((1+e^{\tilde{a}})(1+e^{\tilde{c}})(1+e^{\tilde{\gamma}})(1+e^{\tilde{\phi}_v})\right)\end{aligned} \tag{4.4.19}$$

Remarks: The inverse of $\Sigma_{O|\Omega}$ are evaluated by applying Sherman-Woodbury-Morrison formula (see, e.g., [Har97]), which only requires inverting a matrix with dimension $(n_x^* n_z^* + n_y^*) \times (n_x^* n_z^* + n_y^*)$. Specifically,

$$\begin{aligned}\Sigma_{O|\Omega}^{-1} &= (D_e + B(\theta)\Sigma_{w^*}B^\top(\theta))^{-1}\\ &= D_e^{-1} - D_e^{-1}B(\theta)\left(\Sigma_{w^*}^{-1} + B^\top(\theta)D_e^{-1}B(\theta)\right)^{-1}B^\top(\theta)D_e^{-1}.\end{aligned} \tag{4.4.20}$$

4.5 Predictions

In this section, we do the mapping of y to the locations where the signals Z have been observed through prediction. Section 4.5.1 states the procedure to predict both y and Z at new locations. Then, we derive the prediction of y given Z are known in Section 4.5.2.

4.5.1 Predictions for y and Z at new locations

Assume that there are no observations of y and Z at locations $\tilde{s}_1, ..., \tilde{s}_m$, but there are records of predictors X at those locations.

Let $\tilde{O} = (O^\top(\tilde{s}_1), O^\top(\tilde{s}_2), ..., O^\top(\tilde{s}_m))^\top$. Our goal is to find the conditional distribution of $\tilde{O}$ given all the observations.

We stack the data in a different way so as to separate z and y. Denote by

$$\tilde{O} = \begin{pmatrix} \tilde{Z} \\ \tilde{y} \end{pmatrix}, \tilde{X} = \begin{pmatrix} \tilde{X}_Z & 0 \\ 0 & \tilde{X}_y \end{pmatrix}, \tag{4.5.1}$$

where

$$\tilde{Z} = \begin{pmatrix} Z(\tilde{s}_1) \\ \vdots \\ Z(\tilde{s}_m) \end{pmatrix}, \tilde{y} = \begin{pmatrix} y(\tilde{s}_1) \\ \vdots \\ y(\tilde{s}_m) \end{pmatrix}, \tilde{X}_Z = \begin{pmatrix} X_Z(\tilde{s}_1) \\ \vdots \\ X_Z(\tilde{s}_m) \end{pmatrix}, \tilde{X}_y = \begin{pmatrix} X_y(\tilde{s}_1) \\ \vdots \\ X_y(\tilde{s}_m) \end{pmatrix}. \tag{4.5.2}$$

It is easy to check that

$$\begin{bmatrix} O \\ \tilde{O} \end{bmatrix} |\Omega \sim N\left(\begin{bmatrix} X \\ \tilde{X} \end{bmatrix} b, \begin{bmatrix} \Sigma_{O|\Omega} & C_{O,\tilde{O}|\Omega} \\ C^\top_{O,\tilde{O}|\Omega} & \Sigma_{\tilde{O}|\Omega} \end{bmatrix}\right), \tag{4.5.3}$$

where $\Sigma_{\tilde{O}|\Omega}$ is the covariance matrix of $\tilde{O}$ and $C_{O,\tilde{O}|\Omega}$ is the cross-covariance matrix between O and $\tilde{O}$. Therefore, we can obtain the conditional distribution of $\tilde{O}$ given O and Ω,

$$[\tilde{O}|O,\Omega] \sim N(\mu_{\tilde{O}|O}, \Sigma_{\tilde{O}|O}), \tag{4.5.4}$$

where

$$\begin{aligned} \mu_{\tilde{O}|O} &= \tilde{X}b + C^{\top}_{O,\tilde{O}|\Omega}\Sigma^{-1}_{O|\Omega}(O - Xb) \triangleq (\mu_{\tilde{Z}|O}, \mu_{\tilde{y}|O})^{\top}, \\ \Sigma_{\tilde{O}|O} &= \Sigma_{\tilde{O}|\Omega} - C^{\top}_{O,\tilde{O}|\Omega}\Sigma^{-1}_{O|\Omega}C_{O,\tilde{O}|\Omega} \triangleq \begin{pmatrix} \Sigma_{\tilde{Z}\tilde{Z}} & \Sigma_{\tilde{Z}\tilde{y}} \\ \Sigma^{\top}_{\tilde{Z}\tilde{y}} & \Sigma_{\tilde{y}\tilde{y}} \end{pmatrix}. \end{aligned} \tag{4.5.5}$$

Usually, Bayesian prediction proceeds by sampling from the posterior predictive distribution

$$p(\tilde{O}|O) = \int p(\tilde{O}|O,\Omega)p(\Omega|O)d\Omega \tag{4.5.6}$$

In this case, for each posterior sample of Ω, we draw a corresponding $\tilde{O}$ by (4.5.4).

4.5.2 Predictions for y given Z are observed

Assume that there is no observation of y at locations $\tilde{s}_1, ..., \tilde{s}_m$, but there are records of the signal z and predictors X at those locations. Our goal is to find the conditional distribution of y given all the observations.

By (4.5.4), we can figure out the conditional distribution of $\tilde{y}$ given $\tilde{Z}$, O and Ω,

$$[\tilde{y}|\tilde{Z}, O, \Omega] \sim N(\mu_{\tilde{y}|\cdot}, \Sigma_{\tilde{y}|\cdot}), \tag{4.5.7}$$

where

$$
\begin{aligned}
\mu_{\tilde{y}|\cdot} &= \mu_{\tilde{y}|O} + \Sigma_{\tilde{Z}\tilde{y}}^{\top} \Sigma_{\tilde{Z}\tilde{Z}}^{-1} (\tilde{Z} - \mu_{\tilde{Z}|O}) \\
\Sigma_{\tilde{y}|\cdot} &= \Sigma_{\tilde{y}\tilde{y}} - \Sigma_{\tilde{Z}\tilde{y}}^{\top} \Sigma_{\tilde{Z}\tilde{Z}}^{-1} \Sigma_{\tilde{Z}\tilde{y}}.
\end{aligned} \tag{4.5.8}
$$

The Bayesian prediction proceeds by sampling from the posterior predictive distribution

$$
p(\tilde{y}|\tilde{Z}, O) = \int p(\tilde{y}|\tilde{Z}, O, \Omega) p(\Omega|O, \tilde{Z}) d\Omega \tag{4.5.9}
$$

Since we sample parameters Ω only given the information of O due to computing burden by adding massive size of $\tilde{Z}$, we approximate the posterior predictive distribution as follows,

$$
p(\tilde{y}|\tilde{Z}, O) \approx \int p(\tilde{y}|\tilde{Z}, O, \Omega) p(\Omega|O) d\Omega. \tag{4.5.10}
$$

In this case, for each posterior sample of Ω, we draw a corresponding $\tilde{y}$ by (4.5.7).

4.6 illustrations

We conduct simulation experiments and analyze a large forestry dataset to assess model performance with regard to learning about process parameters and predicting y at new locations. Posterior inference for subsequent analysis were based upon three chains of 30000 iterations (with a burn-in of 5000 iterations). The samplers were programmed in C++ and leveraged Intel's Math Kernel Library's (MKL) threaded BLAS and LAPACK routines for matrix computations. The computations were conducted on a Linux workstation using two Intel Nehalem quad-Xeon processors.

4.6.1 Simulation experiments

We simulate the data on the regular lattices in the domain

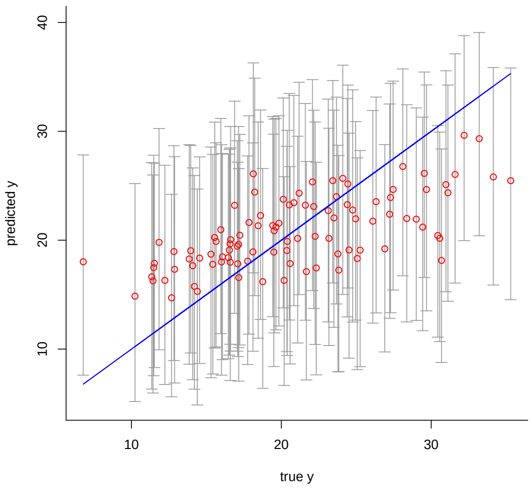
$$\underbrace{[0,4]\times[0,4]}_{\text{location } s}\times\underbrace{[0,5]}_{\text{height } x} \tag{4.6.1}$$

from joint model (4.3.7) with full spatial Gaussian process (GP) where $n = n_z^* = n_y^* = 400$ and $n_x = 50$.

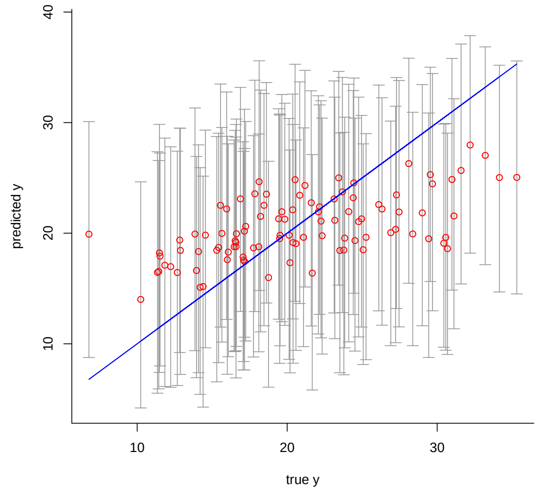
We hold out 25% of data by randomly sampling from the 400 spatial locations. In this study, we choose full knots for y, i.e., $n_y^* = 300$. We fit the following models from the training data: i) the joint model with full knots $n_z^* = 300$; ii) the joint model with $n_z^* = 200$ knots; iii) the joint model with $n_z^* = 100$ knots; iv) the joint model with $n_z^* = 50$ knots. Then, we do predictions for y at the holdout locations for each model.

Parameter estimates and performance metrics for the four models with varying spatial knots for z are provided in Table 4.1. We just listed the estimates of η, α_u and the random effect θ. Larger value of DIC suggest the joint models with fewer knots do not fit the data as well as full knots. Yet, the coefficients α_i which are used to extract information from the signal Z are estimated quite well in each case. The last row in Table 4.1 shows computing times in hours for one chain of 30000 iterations reflecting on the enormous computational gains of predictive process models over full GP model.

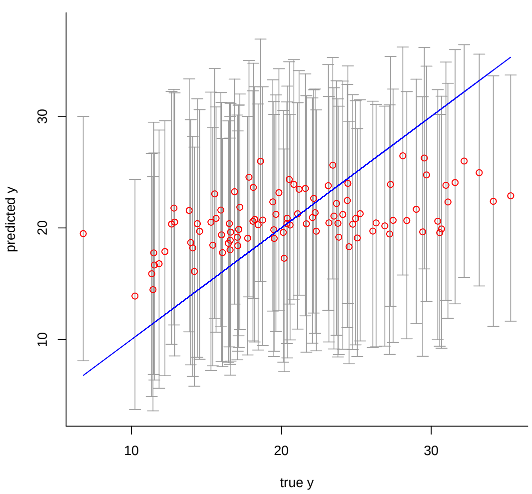
Table 4.1 also indicates that the joint model with full knots for Z has the smallest root mean square prediction error (RMSPE) and smallest mean interval width in terms of predicting y given Z are known. Yet, there is no big difference when we reduce the number of knots for Z. Figure 4.1 shows the 95% credible intervals for predicting 100 holdout y under each model.



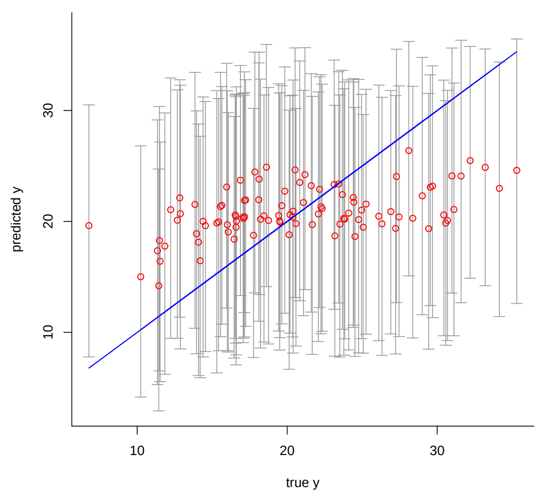
(a) $n_z^* = 300$



(b) $n_z^* = 200$



(c) $n_z^* = 100$



(d) $n_z^* = 50$

Figure 4.1 Predicted y given Z are known. Top left: $n_z^* = 300$ (full knots); top right: $n_z^* = 200$; bottom left: $n_z^* = 100$; bottom right: $n_z^* = 50$. Any red point on the blue line represents the case when the predicted y is equal to the true y.

Table 4.1 Parameter credible intervals, 50% $(2.5\%, 97.5\%)$ and predictive validation. Entries in italics indicate where the true value is missed.

Parameter	True value	Results for the following numbers of knots			
		$n^* = 300$ (Full GP)	$n^* = 200$	$n^* = 100$	$n^* = 50$
η	20	20.45(19.06,21.85)	20.86(19.58,22.11)	20.44(19.28,21.58)	20.5(19.59,21.41)
α_1	-2	-1.83(-2.04,-1.63)	-1.81(-2.05,-1.58)	-1.98(-2.35,-1.57)	-1.82(-2.27,-1.29)
α_2	0	0.12(-0.16,0.37)	0.03(-0.28,0.33)	0.34(-0.14,0.79)	0.3(-0.32,1.07)
α_3	2	2.07(1.77,2.35)	1.91(1.59,2.25)	1.7(1.26,2.16)	1.69(1.13,2.41)
α_4	1	1.07(0.8,1.33)	1.11(0.83,1.39)	1.13(0.72,1.53)	1.27(0.7,1.87)
α_5	5	4.85(4.61,5.09)	4.9(4.64,5.17)	4.97(4.6,5.33)	4.98(4.59,5.37)
σ_u^2	0.2	0.19(0.18,0.2)	0.19(0.18,0.2)	0.2(0.19,0.21)	0.19(0.18,0.2)
a	12	12.91(11.22,15.11)	13.17(11.16,17.44)	12.21(10.64,14.87)	11.92(10.54,13.93)
γ	0.9	0.86(0.78,0.94)	0.83(0.67,0.92)	0.9(0.79,0.98)	0.91(0.82,0.96)
c	5	5.63(4.92,6.43)	*5.73(5.02,6.58)*	5.44(4.69,6.22)	*6.04(5.2,7.04)*
σ_v^2	0.5	0.41(0.1,1.09)	0.29(0.09,0.93)	0.29(0.1,1.08)	0.43(0.1,1.91)
ϕ_v	2	3.63(0.91,7.25)	*7.83(3.59,9.35)*	8.03(1.2,9.48)	*6.08(2.63,9.54)*
p_D		74.96	82.02	77.14	79.95
DIC		28989.64	29226.99	29487.41	29661.21
RMSPE		4.65	4.96	5.26	5.42
95% CI cover %		90	88	88	89
95% CI width		19.70	20.73	21.80	22.54
Time		153.3 h	80.3 h	27.0 h	13.8 h

4.6.2 Forest LiDAR and biomass data analysis

4.6.2.1 Data description and preparation

This dataset was collected on the Penobscot Experimental Forest, Maine. The signals $z(s, x)$ are observed at 26286 locations. At each location, there are 126 measurements equally distributed above ground within $[0, 37.5]$ meters. Among all the locations, there are 451 locations where biomass y is observed. Figure 4.2 below shows roughly how the data look like.

Since the heights of trees are usually smaller than 29.1m at the observed area, there is no signal when the height is above 29.1m at most locations. We first cut the signal at 29.1m. Then, we coarsen the signal within $[0, 29.1]$m by averaging every two consecutive measurements and use them to fit Z. The dimension of signals at each location is set to

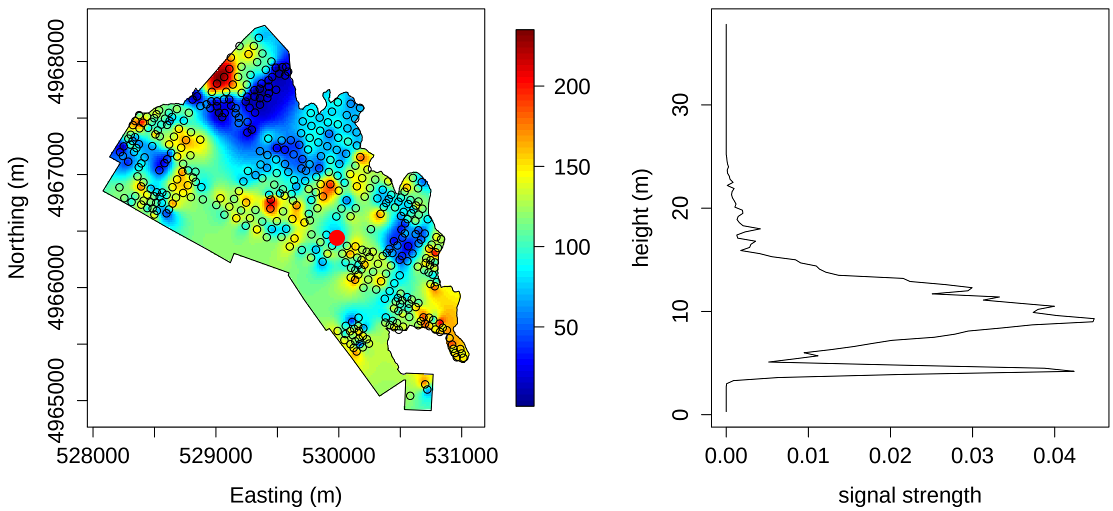


Figure 4.2 Left: Interpolated y, the small points $\circ$ indicate where y's are recorded; Right: Signal Z measured at the big red disc marked on the left graph.

$n_x = 45$. Meanwhile, we sum up the signals above 29.1m at each location and denote by $X_{y,1}(s)$, which is non-zero if the height of tree exceeds 29.1 meters at s . Tall trees usually indicate a large amount of biomass y. So we consider $X_{y,1}(s)$ as the covariate of y across locations. In this study, we do not have other covariates to model Z and y. Hence, we only assume the mean of $Z(s)$ and $y(s)$ are $X_Z^\top(s)\beta = \{\beta_i\}_{i=1}^{n_x}$ and $X_y^\top(s)\eta = \eta_1 + \eta_2 X_{y,1}(s)$ respectively. In addition, for numerical stability, we scale the magnitude of the biomass y and the signal Z, i.e., $y \to y/100$ and $z \to 100z$.

4.6.2.2 Results

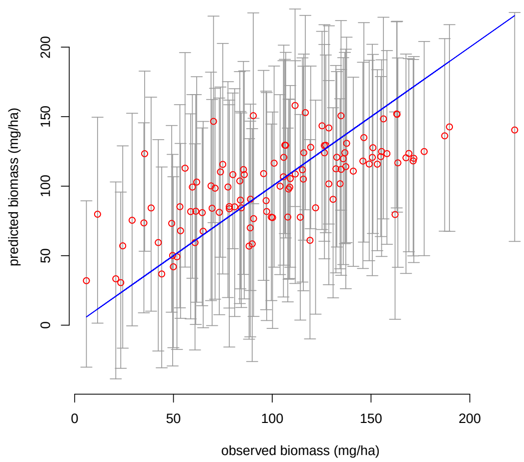
We holdout 25% of data by randomly sampling from the 451 spatial locations. The number of height knots is set to $n_x^* = 5$ and they are evenly distributed across the height. We then choose the same spatial knots for Z and y and fit the following models from the training data: i) the joint model with full knots $n^* = 339$; ii) the joint model with 200 knots; iii) the joint model with 100 knots; iv) the joint model with 50 knots. Finally we do predictions for holdout y and check the prediction performance.

Parameter estimates and performance metrics are provided in Table 4.2. According to our model framework, α_u are used to characterize the relationship between signal and biomass. We see that α_2, α_3 and α_4 are significant. α_2 and α_3 correspond to signals at relatively lower height than α_4 does. Usually, strong signals away from ground indicate big biomass. That's why α_4 is positive while α_2 and α_3 are negative.

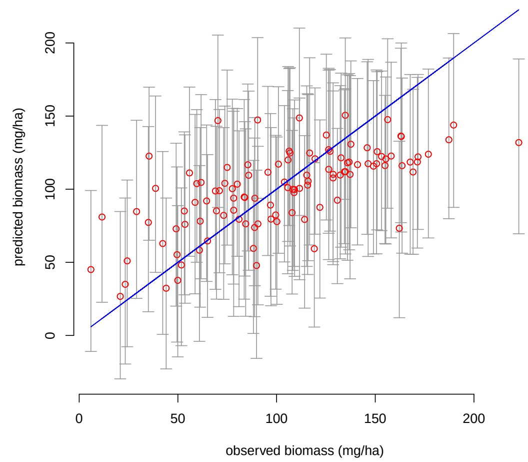
Table 4.2 Parameter credible intervals, 50% (2.5%, 97.5%) and predictive validation.

Parameter	Results for the following numbers of knots			
	$n^* = 339$ (Full GP)	$n^* = 200$	$n^* = 100$	$n^* = 50$
η_1	1.05(0.89,1.23)	1.06(0.72,1.39)	1.05(0.54,1.55)	1.05(0.82,1.29)
η_2	0(-0.04,0.03)	0.01(-0.02,0.03)	0.03(0,0.06)	0.05(0.02,0.07)
α_1	-0.19(-0.61,0.29)	-0.06(-0.23,0.28)	0(-0.17,0.21)	0.06(-0.08,0.19)
α_2	**-0.11(-0.17,-0.05)**	**-0.13(-0.19,-0.07)**	**-0.16(-0.22,-0.1)**	**-0.12(-0.2,-0.03)**
α_3	**-0.09(-0.15,-0.02)**	**-0.11(-0.16,-0.04)**	**-0.14(-0.19,-0.07)**	-0.06(-0.13,0.01)
α_4	**0.09(0.05,0.12)**	**0.06(0.03,0.11)**	**0.05(0,0.1)**	**0.11(0.05,0.17)**
α_5	0.05(-0.09,0.16)	0(-0.12,0.12)	-0.04(-0.17,0.1)	-0.05(-0.17,0.07)
σ_u^2	0.19(0.18,0.21)	0.18(0.17,0.2)	0.18(0.17,0.19)	0.19(0.17,0.2)
a	1.1(1.02,1.18)	1.13(1.05,1.2)	1.19(1.12,1.27)	1.12(1.05,1.19)
γ	0.99(0.97,1)	1(1,1)	0.99(0.99,1)	0.99(0.99,1)
c	8.21(7.56,8.91)	8.18(7.59,8.83)	7.08(6.63,7.63)	5.88(5.51,6.32)
σ_v^2	0.07(0.05,0.09)	0.15(0.11,0.18)	0.19(0.14,0.31)	0.09(0.06,0.12)
ϕ_v	4.38(2.94,6.26)	1.48(1.21,1.83)	0.89(0.88,0.89)	1.86(1.4,2.75)
p_D	86.98	73.90	82.30	80.95
DIC	19111.42	19034.37	18969.00	19253.65
RMSPE for y	0.32	0.33	0.34	0.36
95% CI of y cover %	83	81	80	79
95% CI of y width	1.45	1.17	1.18	1.24
Time	147 h	67.94h	31.94 h	6.37 h

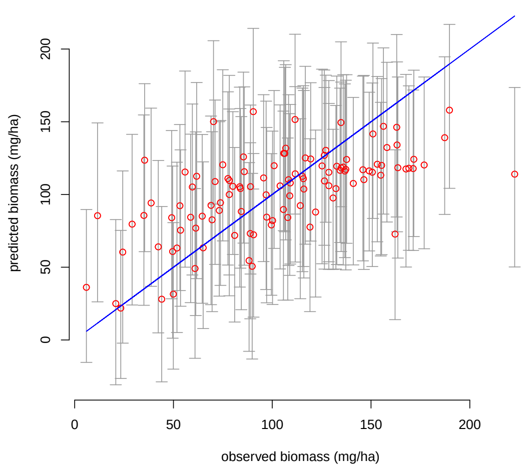
Table 4.2 also indicates that the joint model with full knots has the similar RMSPE and mean CI width as those with reduced number of knots. Figure 4.3 shows the 95% credible intervals for predicting 112 holdout y under each model, from which we see that reducing the number of knots does not affect the prediction of y very much.



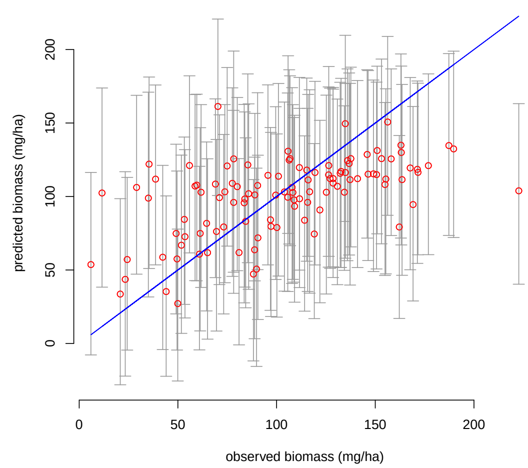
(a) $n^* = 339$



(b) $n^* = 200$



(c) $n^* = 100$



(d) $n^* = 50$

Figure 4.3 Predicted y given Z are known. Top left: $n^* = 339$ (full knots); top right: $n^* = 200$; bottom left: $n^* = 100$; bottom right: $n^* = 50$. Any red point on the blue line represents the case when the predicted biomass is equal to the observed biomass.

BIBLIOGRAPHY

BIBLIOGRAPHY

[AAWN13] A. Awadallah, V. Abbott, R.H. Wynne, and R. Nelson, *Estimating forest canopy height and biophysical parameters using photon-counting laser altimetry*, Proc. 13th International Conference on LiDAR Applications for Assessing Forest Ecosystems (SilviLaser 2013), Oct 2013, pp. 129–136.

[Adl81] Robert J Adler, *The geometry of random fields*, vol. 62, Siam, 1981.

[Adl00] R.J. Adler, *On excursion sets, tube formulas and maxima of random fields*, Ann. Appl. Probab. **10** (2000), 1–74. MR MR1765203 (2001g:60082)

[AGS12] Tatiyana Apanasovich, Marc G. Genton, and Ying Sun, *Cross-covariance functions for multivariate random fields based on latent dimensions*, J. Amer. Statist. Assoc. **107** (2012), no. 497, 180–193. MR MR2949350

[AHV+09] GP Asner, RF Hughes, TA Varga, DE Knapp, and T Kennedy-Bowdoin, *Environmental and biotic controls over aboveground biomass throughout a tropical rain forest*, Ecosystems **12** (2009), no. 2, 261–278.

[Ans06] A.B. Anshin, *On the probability of simultaneous extrema of two Gaussian nonstationary processes*, Theory Probab. Appl. **50** (2006), no. 3, 353–366. MR MR2223210 (2007f:60031)

[AS72] Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, tenth printing*, Dover, New York, 1972.

[AT07] R. Adler and J. Taylor, *Random Fields and Geometry*, New York: Jone Wiley & Sons, 2007.

[ATW10] Robert J Adler, Jonathan E Taylor, and Keith J Worsley, *Applications of random fields and geometry: Foundations and case studies.*

[AW09] Jean-Marc Azaïs and M. Wschebor, *Level Sets and Extrema of Random Procesees and Fields*, Jone Wiley & Sons, Inc., Hoboken, NJ., 2009.

[AZB+10] W Abdalati, HJ Zwally, R Bindschadler, B Csatho, SL Farrell, HA Fricker, R Harding, Dand Kwok, M Lefsky, T Markus, A Marshak, T Neumann, S Palm, B Schutz, B Smith, J Spinhirne, and C Webb, *The icesat-2 laser altimetry mission*, Proceedings of the IEEE **98** (2010), no. 5, 735–751.

[BCG14] Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand, *Hierarchical modeling and analysis for spatial data, 2nd ed.*, CRC press, 2014.

[BGFS08] Sudipto Banerjee, Alan E Gelfand, Andrew O Finley, and Huiyan Sang, *Gaussian predictive process models for large spatial data sets*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **70** (2008), no. 4, 825–848.

[Bil68] Patrick Billingsley, *Convergence of probability measures*, New York: Jone Wiley & Sons, 1968.

[BMF+13] CR Babcock, JA Matney, AO Finley, A Weiskittel, and BD Cook, *Multivariate spatial regression models for predicting individual tree structure variables using lidar data*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **6** (2013), no. 1, SI, 6–14.

[CCN+13] BD Cook, LA Corp, RF Nelson, EM Middleton, DC Morton, JT McCorkel, JG Masek, KJ Ranson, V Ly, and PM Montesano, *Nasa goddards lidar, hyperspectral and thermal (g-liht) airborne imager*, Remote Sensing **5** (2013), no. 8, 4045–4066.

[CD09] Jean-Paul Chiles and Pierre Delfiner, *Geostatistics: modeling spatial uncertainty*, vol. 497, John Wiley & Sons, 2009.

[CH94] A.G. Constantine and Peter Hall, *Characterizing surface smoothness via estimation of effective fractal dimension*, J. Roy. Statist. Soc. Ser. B **56** (1994), no. 1, 97–113. MR MR1257799 (95k:62260)

[CL06] Hock Peng Chan and Tze Leung Lai, *Maxima of asymptotically gaussian random fields and moderate deviation approximations to boundary crossing probabilities of sums of random variables with multidimensional indices*, The Annals of Probability (2006), 80–121.

[Cre93] N.A.C. Cressie, *Statistics for spatial data, 2nd ed.*, New York: Jone Wiley & Sons, 1993.

[CSY00] Huann-Sheng Chen, Douglas G Simpson, and Zhiliang Ying, *Infill asymptotics for a stochastic process model with measurement error*, Statistica Sinica **10** (2000), no. 1, 141–156.

[CW00] Grace Chan and Andrew T.A. Wood, *Increment-based estimators of fractal dimension for two-dimensional surface data*, Statist. Sinica **10** (2000), no. 2, 343–376. MR MR1769748 (2001c:62110)

[CW04] ———, *Estimation of fractal dimension for a class of non-Gaussian stationary processes and fields*, Ann. Statist. **32** (2004), no. 3, 1222–1260. MR MR2065204 (2005i:60062)

[CW11] Noel Cressie and Christopher K Wikle, *Statistics for spatio-temporal data*, John Wiley & Sons, 2011.

[CX14] D. Cheng and Y. Xiao, *Geometry and excursion probability of multivariate Gaussian random fields*, Manuscript (2014).

[DHJ14] KRZYSZTOF DEBICKI, ENKELEJD HASHORVA, and LANPENG JI, *Extremes of a class of non-homogeneous gaussian random fields*, Ann. Probab. (2014).

[DKMR10] K. Debicki, K.M. Kosinski, M. Mandjes, and T. Rolski, *Extremes of multidimensional Gaussian processes*, Stochstic processes and their applications **120** (2010), no. 12, 2289–2301. MR MR2728166 (2011m:60107)

[FBM11] AO Finley, S Banerjee, and DW MacFarlane, *A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas*, Journal ot the American Statistical Association **106** (2011), no. 493, 31–48.

[FSBG09] Andrew O Finley, Huiyan Sang, Sudipto Banerjee, and Alan E Gelfand, *Improving the performance of predictive process modeling for large datasets*, Computational statistics & data analysis **53** (2009), no. 8, 2873–2884.

[GDFG10] A. Gelfand, P. Diggle, M. Fuentes, and P. Guttorp, *Handbook of spatial statistics*, CRC press, 2010.

[GED14] GEDI, *Global ecosystem dynamics investigation lidar*, `http://science.nasa.gov/missions/gedi/`, 2014, Accessed: 1-5-2015.

[GKS10] T. Gneiting, W. Kleiber, and M. Schlather, *Matérn cross-covariance functions for multivariate random fields*, Journal of the American Statistical Association **105** (2010), no. 491, 1167–1177. MR MR2752612 (2012d:62326)

[Gne02] Tilmann Gneiting, *Nonseparable, stationary covariance functions for space–time data*, Journal of the American Statistical Association **97** (2002), no. 458, 590–600. MR MR1941475 (2003k:62283)

[GŠP12] Tilmann Gneiting, Hana Ševčíková, and Donald B. Percival, *Estimators of fractal dimension: assessing the roughness of time series and spatial data*, Statist. Sci. **27** (2012), no. 2, 247–277. MR 2963995

[Har97] David A Harville, *Matrix algebra from a statistician's perspective*, vol. 157, Springer, 1997.

[HJ14] E. Hashorva and L. Ji, *On the Joint Extremes of Two Correlated Fractional Brownian Motions*, arXiv:1309.4981 [math.PR] (2014).

[HW71] David Lee Hanson and Farroll Tim Wright, *A bound on tail probabilities for quadratic forms in independent random variables*, The Annals of Mathematical Statistics (1971), 1079–1083.

[HW93] Peter Hall and Andrew Wood, *On the performance of box-counting estimators of fractal dimension*, Biometrika **80** (1993), no. 1, 246–252. MR MR1225230 (94m:62108)

[ICE15] ICESat-2, *Ice, cloud, and land elevation satellite-2*, `http://icesat.gsfc.nasa.gov/`, 2015, Accessed: 1-5-2015.

[IDUA13] IA Iqbal, J Dash, S Ullah, and G Ahmad, *A novel approach to estimate canopy height using icesat/glas data: A case study in the new forest national park, uk*, International Jorunal of Applied Earth Observation and Geoinformation **23** (2013), 109–118.

[KN12] William Kleiber and Douglas Nychka, *Nonstationary modeling for multivariate spatial processes*, Journal of Multivariate Analysis **112** (2012), 76–91.

[KW95] John T. Kent and Andrew T. A. Wood, *Estimating the fractal dimension of a locally self-similar Gaussian process using increments*, Statistics Research Report SRR 034-95. (1995), Centre for Mathematics and Its Applications, Australian National University, Canberra.

[KW97] John T. Kent and T.A. Wood, *Estimating the fractal dimension of a locally self-similar Gaussian process by using increments*, J. Roy. Statist. Soc. Ser. B **59** (1997), no. 3, 679–699. MR MR1452033 (99a:62136)

[LP00] A. Ladneva and V.I. Piterbarg, *On double extremes of Gaussian statioinary processes*, EURANDOM Techinal report 2000-027 (2000), 1–18, Available from: `http://www.eurandom.tue.nl/reports/2000/027-report.pdf`[Accessed Mar.19th 2014].

[Mey00] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphis, 2000.

[MMT11] JD Muss, DJ Mladenoff, and PA Townsend, *A pseudo-waveform technique to assess forest structure using discrete lidar data*, Remote Sensing of Environment **115** (2011), no. 3, 824–835.

[MN07] J.R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics, revised ed.*, John Wiley & Sons, 2007.

[Næs11] E Næsset, *Estimating above-ground biomass in young forests with airborne laser scanning*, International Journal of Remote Sensing **32** (2011), no. 2, 473–501.

[NNR+13] CSR Neigh, RF Nelson, KJ Ranson, HA Margolis, PM Montesano, G Sun, V Kharuk, E Næsset, MA Wulder, and H Andersen, *Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne lidar*, Remote Sensing of Environment **137** (2013), 274–287.

[NSY08] Y. Nardi, D. O. Siegmund, and B. Yakir, *The distribution of maxima of approximately Gaussian random fields*, Ann. Statist. **36** (2008), 1375–1403. MR MR2418661 (2009e:60112)

[Pic69] James Pickands, *Upcrossing probabilities for stationary gaussian processes*, Transactions of the American Mathematical Society **145** (1969), 51–73.

[Pit96] V. Piterbarg, *Asymptotic methods in the theory of Gaussian processes and fields*, American Mathematical Society, 1996.

[PS05] V.I. Piterbarg and B. Stamatovich, *Rough asymptotics of the probability of simultaneous high extrema of two Gaussian processes: the dual action functional*, Russ. Math. Surv. **60** (2005), no. 1, 167–168. MR MR2145669 (2006a:60064)

[RB13] Qian Ren and Sudipto Banerjee, *Hierarchical factor models for large spatially misaligned data: a low-rank predictive process approach*, Biometrics **69** (2013), no. 1, 19–30. MR 3058048

[Ste99] Michael Stein, *Interpolation of spatial data: some theory for kriging*, Springer, 1999.

[Wac03] H. Wackernagel, *Multivariate geostatistics: an introduction with applications, 2nd ed.*, Springer, 2003.

[Yag87] AM Yaglom, *Correlation theory of stationary and related random functions: Vol.: 1: Basic results*, Springer-Verlag, 1987.

[Yak13] B. Yakir, *Extremes in Random Fields*, Higher Education Press, 2013.

[ZS02] Zhengyuan Zhu and Michael L. Stein, *Parameter estimation for fractional Brownian surfaces*, Statist. Sinica **12** (2002), no. 3, 863–883. MR MR1929968