Ph.D. DISSERTATION

# Spectral tail parameter estimation under fixed domain asymptotics

## 고정 도메인에서의 스펙트럴 꼬리 행동 모수 추정

FEBRUARY 2023

DEPARTMENT OF STATISTICS
COLLEGE OF NATURAL SCIENCE
SEOUL NATIONAL UNIVERSITY

Joonho Shin

Ph.D. DISSERTATION

# Spectral tail parameter estimation under fixed domain asymptotics

## 고정 도메인에서의 스펙트럴 꼬리 행동 모수 추정

FEBRUARY 2023

DEPARTMENT OF STATISTICS
COLLEGE OF NATURAL SCIENCE
SEOUL NATIONAL UNIVERSITY

Joonho Shin

Spectral tail parameter estimation
under fixed domain asymptotics

# 고정 도메인에서의 스펙트럴 꼬리 행동 모수 추정

지도교수 임 채 영

이 논문을 이학박사 학위논문으로 제출함

2022 년 10 월

서울대학교 대학원

통계학과

신 준 호

Joonho Shin의 이학박사 학위논문을 인준함

2022 년 12 월

| | | |
|---|---|---|
| 위 원 장 | 오희석 | (인) |
| 부위원장 | 임채영 | (인) |
| 위 원 | 이재용 | (인) |
| 위 원 | 신예은 | (인) |
| 위 원 | 박재우 | (인) |

# Abstract

When it comes to fixed domain asymptotics for stationary Gaussian random fields, there are many results emphasizing the role of tail behavior of spectral densities. Examples are smoothness of random fields, asymptotically optimal kriging under misspecified spectral density and equivalence of Gaussian measures. Assuming parametric tail behavior structure on spectral density, this dissertation aims to estimate parameters therein consistently under fixed domain asymptotics. Specifically, we focus on the data collected on a regular lattice in a fixed bounded subset of $\mathbb{R}^d$ and develop periodogram-based spectral domain approach to construct objective functions which resemble Whittle likelihood. Smoothed periodogram emerges as an important statistics during the construction, thus we first proved consistency and asymptotic normality of smoothed periodogram under fixed domain asymptotics. Next, based on two types of smoothed peridograms, tapered periodogram and smoothed periodogram with a compactly supported kernel, we construct a Whittle type objective function whose minimizer becomes an estimator for spectral tail behavior parameters. Consistency and asymptotic order of our estimators are derived with asymptotic normality for some cases. As a byproduct, our result enables statistical inference for the estimated parameters. Simulation experiments are included to support our theoretical results. As real world applications, we analyze sea ice profiles data and monthly maximum temperature data.

**Keywords**: Fixed domain asymptotics, Gaussian random field, Periodogram, Spectral density

**Student Number**: 2016-20269

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Throughout this dissertation, we will consider the data observed on a $d$-dimensional rectangular lattice in a bounded region. When the data is collected from a real-valued mean zero stationary Gaussian process, many statistical procedures begins with estimating the underlying true measure or equivalently the autocovariance structure. Using the relation between autocovariance functions and spectral densities, we can instead estimate spectral density to get covariance structure of the process. The structure is often fully or partially parametrized, in which case estimating these parameters is of great interest. For the most cases, deriving finite sample properties of parameter estimators is difficult and mathematically intractable in general so that asymptotic properties of these parameters are usually investigated. Asymptotic framework suitable to describe densely observed data over a bounded region is called the fixed domain asymptotics.

We will focus on estimating the tail behavior of $f$, which means the behavior of $f(\boldsymbol{\omega})$ as $|\boldsymbol{\omega}| \to \infty$. We will consider a family of tail behavior of spectral density which can be expressed as follows:

$$f(\boldsymbol{\omega}) \sim c|\boldsymbol{\omega}|^{-\alpha} h\left(\frac{\boldsymbol{\omega}}{|\boldsymbol{\omega}|}; \alpha, \boldsymbol{\beta}\right) \tag{1.1}$$

as $|\boldsymbol{\omega}| \to \infty$, where $c > 0$, $\alpha > d$, $h$ is a bounded function on a unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$ and $a(\boldsymbol{\omega}) \sim b(\boldsymbol{\omega})$ means that $a(\boldsymbol{\omega})/b(\boldsymbol{\omega}) \to 1$ as $|\boldsymbol{\omega}| \to \infty$. We call $(c, \alpha, \boldsymbol{\beta}^T)^T$ the spectral tail parameter, or simply the tail parameter and denote it as $\boldsymbol{\theta}$. The main goal is to propose a method to estimate $\boldsymbol{\theta}$ and prove fixed domain asymptotic properties of the estimator.

Model (1.1) covers wide range of covariance models which are frequently used for random field modeling. For example, the isotropic Matérn covariance function, which is given as

$$C(\mathbf{s}; \sigma^2, \nu, a) = \frac{\sigma^2 (a|\mathbf{s}|)^\nu}{\Gamma(\nu) 2^{\nu-1}} \mathcal{K}_\nu(a|\mathbf{s}|), \quad \mathbf{s} \in \mathbb{R}^d,$$

where $\sigma^2$ and $a > 0$ are the variance and scale parameters, $\nu > 0$ is the smoothness parameter and $\mathcal{K}_\nu$ is the modified Bessel function of the second kind, has the spectral density

$$f(\boldsymbol{\omega}; \sigma^2, \nu, a) = \frac{\sigma^2 a^{2\nu} \Gamma(\nu + d/2)}{\pi^{d/2} \Gamma(\nu) (a^2 + |\boldsymbol{\omega}|^2)^{\nu + d/2}}, \quad \boldsymbol{\omega} \in \mathbb{R}^d.$$

Observing $f(\boldsymbol{\omega}; \sigma^2, \nu, a) \sim \frac{\sigma^2 a^{2\nu} \Gamma(\nu + d/2)}{\pi^{d/2} \Gamma(\nu)} |\boldsymbol{\omega}|^{-2\nu - d}$, the isotropic Matérn model satisfies (1.1) by reparametrizing $(c, \alpha, a) = \left(\frac{\sigma^2 a^{2\nu} \Gamma(\nu + d/2)}{\pi^{d/2} \Gamma(\nu)}, 2\nu + d, a\right)$. When geometric anisotropy is introduced on the Matérn covariance structure so that the covariance function becomes $C(A\mathbf{s}; \sigma^2, \nu, a, A)$ with an anisotropy matrix $A$,

its spectral density can be written simply by replacing $|\boldsymbol{\omega}|$ into $|A^{-T}\boldsymbol{\omega}|$ from that of the isotropic Matérn spectral density. Adopting previous reparametrization, it is easy to see that $f(\boldsymbol{\omega}; c, \alpha, A, a) \sim c|A^{-T}\boldsymbol{\omega}|^{-\alpha}$. By setting $h(\boldsymbol{\omega}/|\boldsymbol{\omega}|; \alpha, A) = (|A^{-T}\boldsymbol{\omega}|/|\boldsymbol{\omega}|)^{-\alpha}$, we get the tail behavior of the form in (1.1).

Note that only the tail of spectral density, not the whole structure, is modeled in (1.1). There are some theoretical results from which we can notice that spectral tail behavior matters under fixed domain asymptotics. One is about the equivalence of Gaussian measures and its relation with parameter estimability. Look at the following which is Theorem A.1. of Stein (2004):

> Suppose that $\mathbb{P}_0$ and $\mathbb{P}_1$ are two Gaussian measures that makes $Z(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d$ a mean zero stationary Gaussian random field with spectral densities $f_0$ and $f_1$. If, for some $\alpha > d$, $f_0(\boldsymbol{\omega})|\boldsymbol{\omega}|^\alpha$ is bounded away from 0 and $\infty$ as $|\boldsymbol{\omega}| \to \infty$ and, for some finite $R$,
>
> $$\int_{|\boldsymbol{\omega}|>R} \left\{ \frac{f_1(\boldsymbol{\omega}) - f_0(\boldsymbol{\omega})}{f_0(\boldsymbol{\omega})} \right\}^2 \mathbf{d}\boldsymbol{\omega} < \infty, \tag{1.2}$$
>
> then $\mathbb{P}_0$ and $\mathbb{P}_1$ are equivalent on the path of $Z(\mathbf{s}), \mathbf{s} \in \mathcal{D}$, for any bounded subset $\mathcal{D} \subset \mathbb{R}^d$.

This theorem serves an important role in Zhang (2004) to prove that in the isotropic Matérn model it is impossible to estimate $\sigma^2$ and $a$ separately when $\nu$ is known and $d \leq 3$ based on any sample from a bounded region. The proof is essentially to show that any two spectral densities $f_0$ and $f_1$ contained in $\{f(\boldsymbol{\omega}; \sigma^2, \nu, a) : \sigma^2 a^{2\nu} = C\}$ for some $C > 0$ with a fixed $\nu$ meet the criterion (1.2), which holds because the ratio $f_1(\boldsymbol{\omega})/f_0(\boldsymbol{\omega})$ converges to 1 fast enough. This means that parameters which make different spectral densities

close enough on their tails cannot be consistently estimated under fixed domain asymptotics, from which we can glimpse the relation between spectral tail behavior and parameter estimability.

Another is related to an interpolation problem. Specifically, Theorem 12 in page 136 of Stein (1999) implies that if $f_0 \in \mathcal{Q}^d$ and $f_1(\boldsymbol{\omega})/f_0(\boldsymbol{\omega}) \to c$ as $|\boldsymbol{\omega}| \to \infty$ for some positive $c$, then the linear interpolation obtained by using $f_1$ is asymptotically optimal if the true spectral density is $f_0$ when the observation becomes dense on a bounded region in $\mathbb{R}^d$. Here $\mathcal{Q}^d$ is the set of functions $f : \mathbb{R}^d \to \mathbb{R}$ such that $f(\boldsymbol{\omega})/|\varphi(\boldsymbol{\omega})|^2$ is bounded and away from zero as $|\boldsymbol{\omega}| \to \infty$ for some function $\varphi$ which is the Fourier transform of a square integrable function with a bounded support. This suggests us the importance of spectral tail behavior in that we only need to specify tail behavior of spectral density when our purpose is prediction.

The spectral tail behavior is also closely related to the roughness or smoothness of random fields. The fractal or Hausdorff dimension is a measure for such properties. For instance, suppose that $Z(\mathbf{s})$ is a stationary Gaussian random field on $\mathbb{R}^d$ whose variogram $\gamma_2(\mathbf{s}) = \frac{1}{2}\mathbb{E}(Z(\mathbf{x}+\mathbf{s}) - Z(\mathbf{x}))^2 = C(\mathbf{0}) - C(\mathbf{s})$ satisfies

$$\gamma_2(\mathbf{s}) = |c_2\mathbf{s}|^\alpha + O(|\mathbf{s}|^{\alpha+\beta}) \quad \text{as } |\mathbf{s}| \to 0,$$

where $\alpha \in (0, 2], \beta \geq 0$ and $c_2 > 0$. Then it is known that the graph of $Z$ has the fractal dimension $D = d + 1 - \frac{\alpha}{2}$. The graph becomes smoother when $D$ takes values close to $d$, and becomes rougher as it gets close to $d+1$ (Gneiting et al., 2012). The behavior of the variogram around the origin, or equivalently that of the covariance function around the origin, is closely related to the decay rate

of the spectral measure on its tail, which is a result of Abelian and Tauberian theorems (Stein, 1999, Section 2.8). Since spectral density is just a Radon-Nikodym derivative of a spectral measure, relation between a decay rate of a spectral density and smoothness of a random field is revealed.

In order to estimate spectral density, periodogram is the most classical yet widely used statistics. Periodogram is a nonparametric estimator for spectral density originated from time series analysis. It is well known that for a stationary time series with a fixed time interval, periodogram is an asymptotically unbiased estimator for the spectral density (e.g. Brillinger, 2001). Its variance however does not vanish so that the periodogram is not consistent. Decorrelation effect of a discrete Fourier transform used in the evaluation of periodogram provides a remedy by smoothing periodogram to lower the variance. Smoothed peridogram itself can be regarded as a nonparametric estimate for spectral density, and used to estimate the parameters in spectral density as well. Analogue for random fields, which is an extension toward $d \geq 2$, can be found in literature such as Guyon (1982) and Heyde and Gay (1993). It is important to mention that an asymptotic framework underlying these results are called increasing domain asymptotics, which is an assumption that the distance between neighboring observations is fixed while the size of the observation domain increases.

Natural desire would be to extend the methodologies in the increasing domain setting toward fixed domain perspective. However, it incurs extremely cumbersome calculations for theoretical investigation in general. To the best of our knowledge, Stein (1995) was the first successful attempt that investigated theoretical aspects of periodogram for random fields under fixed domain

asymptotic circumstances. Emphasizing the tail behavior of spectral density, Stein (1995) assumed $f(\boldsymbol{\omega}) \sim c|\boldsymbol{\omega}|^{-\alpha}$ and investigated periodogram when the data is observed on a lattice lying in $[0,1]^d$ with a neighboring distance $\delta$ which tends to 0. To avoid blow-up tendency of the aliased spectral density around the origin, Stein (1995) proposed a filter called discrete Laplacian operator and proved that spatial periodogram achieved from the data filtered with this filter has asymptotic unbiasedness and uncorrelatedness under fixed domain asymptotics. Based on these theories, estimating $c$ and $\alpha$ through mimicking Whittle's likelihood approximation had been attempted in its discussion section. Although simulation results were depicted, the estimation procedure had not been theoretically investigated.

Inspired by Stein (1995), we define the smoothed spatial periodogram and investigate its asymptotic properties under fixed domain asymptotics when the underlying spectral density satisfies (1.1). Our smoothed periodogram is defined by integrating periodogram with a kernel function. We then define an objective function using smoothed periodograms whose form can be seen as a locally smoothed version of Whittle's approximated likelihood. By minimizing the objective function, we prove that the minimizer becomes a consistent estimator for the spectral tail parameter $\boldsymbol{\theta} = (c, \alpha, \boldsymbol{\beta}^T)^T$ in (1.1). The order of convergence for our estimator is also derived as well as asymptotic distribution for some cases.

# Chapter 2

# Literature Reviews

## 2.1 Studies on Periodograms for Random Fields

As in our method, most spectral domain approaches are based on periodograms. Periodograms for random fields are studied in various literature. Guyon (1982) studied periodogram of a stationary random fields on a $d$-dimensional lattice under increasing domain perspective and discovered that there is a bias in periodogram called the edge effect which becomes significant when $d \geq 2$. He suggested to use an unbiased periodogram to remove the edge effect. Heyde and Gay (1993) derived asymptotic distribution of smoothed periodogram for random fields under increasing domain setting where the periodogram therein is unbiased as in Guyon (1982). As mentioned before, Stein (1995) investigated spatial periodograms under fixed domain asymptotic framework. Lim and Stein (2008) generalized the result of Stein (1995) toward multivariate random fields

and derived asymptotic normality of locally smoothed spatial periodogram matrix.

Some trials have been made to apply periodogram to irregularly spaced data. Matsuda and Yajima (2009) constructed irregularly spaced observation on $\mathbb{R}^d$ via sampling location randomization and considered mixed asymptotic regime. They evaluated periodogram from irregularly spaced stationary random fields and showed that smoothed periodogram serves as a nonparametric estimate for spectral density. Rao (2018) studied periodogram based statistics for randomly spaced data under increasing domain asymptotics and fixed domain asymptotics. Guinness (2019) used periodic embedding to estimate spectral density of a random field on a $d$-dimensional lattice from incomplete gridded data.

Compared to the aforementioned works, our result for smoothed peridogram can be seen as a fixed domain asymptotic version of Heyde and Gay (1993) and is different from Stein (1995) and Lim and Stein (2008) since our smoothed periodogram is defined by global integration over frequency domain where those in two works are defined by weighted local summation on a shrinking set in a frequency domain.

## 2.2 Parameter Estimation using Spectral Analysis

Parameter estimation through spectral domain approach usually involves Whittle's idea of likelihood approximation and its variants. Guyon (1982) constructed negative Whittle likelihood using smoothed unbiased periodogram and prove consistency and asymptotic order of its minimizer. Robinson (1995) considered

the Gaussian time series model with spectral density $f$ behaves like $f(\omega) \sim c|\omega|^{-\alpha}$ as $|\omega| \to 0$ for $c > 0$ and $\alpha \in (-1, 1)$. Under increasing domain setting, he showed that the minimizer of partially discretized negative Whittle's approximated likelihood becomes a consistent estimator for $\alpha$ with asymptotic normality. Stein (1995) is the work on fixed domain asymptotics when the spectral density $f$ of Gaussian random field reveals the tail behavior $f(\boldsymbol{\omega}) \sim c|\boldsymbol{\omega}|^{-\alpha}$ as $|\boldsymbol{\omega}| \to \infty$ for $c > 0$ and $\alpha > d$. By constructing discretized Whittle likelihood based on his spatial periodogram theory, he showed simulation results for estimating $c$ and $\alpha$.

Sharing the same spectral density assumption as in Stein (1995), Wu et al. (2013) and Wu and Lim (2016) proposed estimators for $c$ and $\alpha$ under fixed domain asymptotic regime. Their estimation were based on the negative local Whittle likelihood whose theories are based on periodogram properties demonstrated in Lim and Stein (2008). Wu et al. (2013) proposed a consistent estimator for $c$ when $\alpha$ is known and a consistent estimator for $\alpha$ when $c$ is an arbitrary fixed value, and Wu and Lim (2016) more focused on the estimation of the smoothness parameter $\alpha$ when $c$ is unknown and reduced the bias of the estimator.

For irregularly spaced data, Fuentes (2007) approximated the likelihood via Whittle likelihood when the irregularity is considered as missing values on a regular lattice in $\mathbb{R}^d$. Matsuda and Yajima (2009) showed that the minimizer of a local Whittle likelihood constructed with the smoothed periodogram evaluated from their irregular data periodogram serves as a consistent parameter estimator with asymptotic normality under mixed domain asymptotics. Investigation

on the Whittle likelihood itself to reduce the bias can be found in Sykulski et al. (2019) and Rao and Yang (2021), both of which are targeted to time series data. Instead of removing bias from periodogram as in Guyon (1982), Sykulski et al. (2019) removed the bias by replacing the spectral density in Whittle likelihood into the actual expected value of periodogram. Rao and Yang (2021) compared Whittle likelihood with the original Gaussian likelihood to get an approximation error and proposed a new version of the Whittle likelihood which can yield better an estimator with satisfactory finite sample properties.

Our work is related to Robinson (1995); Wu et al. (2013); Wu and Lim (2016) in that parameters in semiparametric models for spectral density has been estimated. Especially, our work is close to Wu et al. (2013) and Wu and Lim (2016) because these works also deal with the spectral tail parameter estimation. However, the spectral tail behavior model in our work incorporates much broader and general classes of spectral densities compared to them. Moreover, if we assume $f(\boldsymbol{\omega}) \sim c|\boldsymbol{\omega}|^{-\alpha}$ as $|\boldsymbol{\omega}| \to \infty$ which is a special case of (1.1), the method we propose can simultaneously estimate $c$ and $\alpha$ consistently with asymptotic normality which had not been possible in Wu et al. (2013) and Wu and Lim (2016). Additionally, our idea to remove bias in the stage of constructing an objective function is similar to that in Sykulski et al. (2019).

## 2.3 Parameter Estimation under Fixed Domain Asymptotics

Under fixed domain perspective, spatial domain approaches to estimate parameters in a covariance function are also actively studied. Ying (1991) shows that when $Z(s), s \in [0,1]$ is Gaussian with an exponential covariance structure $C(s;\sigma^2,a) = \sigma^2 e^{-a|s|}$, we cannot estimate both $\sigma^2$ and $a$ consistently. Note that this is the special case of Matérn when $\nu = 0.5$. Instead, it is shown that $\sigma^2 a$, called the microergodic parameter in literature, can be consistently estimated through MLE methods.

Ying (1993) shows that when $Z(\mathbf{s}), \mathbf{s} \in [0,1]^d$ is Gaussian and the covariance structure is multiplicative exponential given as $C(\mathbf{s};\sigma^2,a_1,\cdots,a_d) = \sigma^2 e^{-\sum_{j=1}^{d} a_1|s_j|}$ and $d \geq 2$, all the parameters $\sigma^2$ and $a_1,\cdots,a_d$ are consistently estimable through the MLE based on complete lattice observations. Zhang (2004) proves that when $Z(\mathbf{s}), \mathbf{s} \in \mathcal{D}$ for a bounded subset $\mathcal{D} \subset \mathbb{R}^d$ with a Matérn covariance structure, we cannot estimate both $\sigma^2$ and $a$ consistently when the smoothness parameter $\nu$ is known. However the microergodic parameter $\sigma^2 a^{2\nu}$ can be estimated consistently through the MLE by fixing $a$ as any value, say $a_1$, and optimizing the likelihood with respect to $\sigma^2$. This result holds when the observations at different stages are nested, $\nu$ is known and $d \leq 3$. Kaufman et al. (2008) studied the fixed domain asymptotics of a tapered MLE, which pursues computational efficiency in inverting a covariance matrix, for $\sigma^2 a^{2\nu}$ when the covariance structure is Matérn with a known $\nu$ and $d \leq 3$. They proved the strong consistency of their estimator. Du et al. (2009) further

derived asymptotic distribution of a tapered MLE in Kaufman et al. (2008) when $d = 1$, and Wang and Loh (2011) extended this result to $d \leq 3$.

For the general $d$, Anderes (2010) showed that $\sigma^2 a^{2\nu}$ and $A$ can be both consistently estimated using squared increments when the underlying covariance model is anisotropic Matérn with an anisotropy matrix $A$ and a known $\nu$. Moreover, the work also showed that $\sigma^2$ and $a$ can also be estimated separately when $d \geq 5$. Note that this work assumed lattice observation as in our work. Bevilacqua et al. (2019) assumed their covariance function belongs to a generalized Wendland class. This model contains four parameters, $\mu, \kappa, \beta$ and $\sigma^2$. Here $\kappa$ can be seen as a smoothness parameter whose role is similar to $\nu$ in the Matérn class. By assuming $\mu$ and $\kappa$ are known, they showed that $\sigma^2 \beta^{-1-2\kappa}$ can be consistently estimated under fixed domain asymptotics when $d \leq 3$. While all the aforementioned results treated the smoothness parameter known, Gneiting et al. (2012) provides a comprehensive review for estimating a smoothness parameter. Most of the methods presented in it are about fractal dimension of time series, which is related to the smoothness parameter when $d = 1$. Some methods applicable for spatial data with $d = 2$ are also introduced in its Section 4. For the simultaneous estimation of a microergodic parameter and a smoothness parameter of the Matérn class, Loh et al. (2021) recently proposed an estimator for $\nu$ and then estimated $\sigma^2 a^{2\nu}$ based on the estimated $\nu$ so that both are consistent under the fixed domain asymptotics with various sampling designs when $d \leq 3$.

Since our method can also be applied to a parametric model, we can compare our estimator to those in the literature. When we apply our method to the

Matérn model, for instance, what we can get is a jointly consistent estimator for $\sigma^2 a^{2\nu}$ and $\nu$ as in Loh et al. (2021). One difference is on the sampling scheme, which is a regular lattice, while various irregular lattice designs are covered in Loh et al. (2021). Another is while the consistency in Loh et al. (2021) is verified $d \leq 3$, our method is proved to be valid for general $d$.

# Chapter 3

# Preliminaries

## 3.1 Prewhitening Random Fields and Periodogram

Throughout this work, we suppose that $Z$ is a real-valued stationary Gaussian random field on $\mathbb{R}^d$ with spectral density $f(\boldsymbol{\omega}), \boldsymbol{\omega} \in \mathbb{R}^d$. We assume that our observation is from $\delta\mathbb{Z}^d$ so that the distance between neighboring observations in each direction is $\delta$. A fixed domain asymptotic regime is then endowed by taking $\delta \downarrow 0$. Note that $\{Z(\delta\mathbf{J}), \mathbf{J} \in \mathbb{Z}^d\}$ can be seen as a lattice process on $\mathbb{Z}^d$ whose spectral density $\bar{f}_\delta(\boldsymbol{\omega}), \boldsymbol{\omega} \in \mathbb{T}^d$ is given as

$$\bar{f}_\delta(\boldsymbol{\omega}) = \frac{1}{\delta^d} \sum_{\mathbf{Q} \in \mathbb{Z}^d} f\left(\frac{\boldsymbol{\omega} + 2\pi\mathbf{Q}}{\delta}\right).$$

Here $\mathbb{T}$ is a torus $\mathbb{R}/2\pi\mathbb{Z} = [-\pi, \pi]$. On $\mathbb{T}$, $-\pi$ and $\pi$ are identified to equip a periodic structure. Hence functions defined on a $d$-dimensional torus $\mathbb{T}^d$ which will constantly appear in this work should be understood to possess periodicity.

Unfortunately, $\bar{f}_\delta$ is changing under the fixed domain asymptotics as $\delta \downarrow 0$

and it typically shows a blow-up tendency around the origin. For instance, suppose $f(\boldsymbol{\omega}) \sim c|\boldsymbol{\omega}|^{-\alpha}$ as $|\boldsymbol{\omega}| \to \infty$. For $f$ to be integrable, we need $\alpha > d$. We then have $\delta^{d-\alpha}\bar{f}_\delta(\boldsymbol{\omega}) \to c\sum_{\mathbf{Q}\in\mathbb{Z}^d}|\boldsymbol{\omega}+2\pi\mathbf{Q}|^{-\alpha}$ as $\delta \downarrow 0$ for each $\boldsymbol{\omega} \neq \mathbf{0}$. The limit function is clearly not integrable since it explodes like $c|\boldsymbol{\omega}|^{-\alpha}$ around the origin. For further explanation, see Stein (1995). As a remedy to suppress this explosive behavior, Stein (1995) suggested to prewhiten the data using the *discrete Laplacian operator*, iteratively, if it is needed. Since we also take this idea into account, we briefly review this operator. Define the lattice process $Y_\delta(\mathbf{J}), \mathbf{J} \in \mathbb{Z}^d$ as $Y_\delta(\mathbf{J}) = Z(\delta\mathbf{J})$. The process after applying the discrete Laplacian operator $\tau$-times is defined by the iterative relation

$$Y_\delta^\tau(\mathbf{J}) = \sum_{j=1}^{d}\left\{Y_\delta^{\tau-1}(\mathbf{J}+\mathbf{e}_j) - 2Y_\delta^{\tau-1}(\mathbf{J}) + Y_\delta^{\tau-1}(\mathbf{J}-\mathbf{e}_j)\right\}, \quad \tau \in \mathbb{N}$$

with $Y_\delta^0(\mathbf{J}) = Y_\delta(\mathbf{J})$, where $\mathbf{e}_j$ is the unit vector along the $j$th coordinate axis. Then the spectral density of $Y_\delta^\tau(\mathbf{J})$, denoted by $\bar{f}_\delta^\tau(\boldsymbol{\omega})$, is given as

$$\bar{f}_\delta^\tau(\boldsymbol{\omega}) = \frac{1}{\delta^d}\left\{\sum_{j=1}^{d} 4\sin^2\left(\frac{\omega_j}{2}\right)\right\}^{2\tau} \sum_{\mathbf{Q}\in\mathbb{Z}^d} f\left(\frac{\boldsymbol{\omega}+2\pi\mathbf{Q}}{\delta}\right), \quad \boldsymbol{\omega} \in [-\pi,\pi]^d.$$

By taking $\tau$ such that $4\tau - \alpha > 0$ under the the tail behavior assumption, we can successfully suppress the explosive behavior as we desired.

Now, consider that we have the filtered data $\{Y_\delta^\tau(\mathbf{J}) : \mathbf{J} \in \mathbb{Z}_N^d\}$ where $\mathbb{Z}_N = \{0, 1, \cdots, N-1\}$. Since this data is on a rectangular lattice comprised of $N^d$-data with neighboring distance $\delta$ along each coordinate axis, we can still impose the fixed domain asymptotic circumstance by letting $N\delta$ fixed. Hereafter, we set $N\delta = 1$ without loss of generality, since the whole theory can be adapted straightforwardly for the case when $N\delta = b \neq 1$. A nonparametric

estimator of the spectral density called fixed domain periodogram, or simply periodogram evaluated from this form of data is defined as

$$\mathcal{I}^{\mathcal{T}}_{N,\delta}(\boldsymbol{\omega}) = \frac{1}{(2\pi N)^d}\left|\sum_{\mathbf{J}\in\mathbb{Z}^d_N} Y^{\mathcal{T}}_{\delta}(\mathbf{J})\exp^{-i\langle\boldsymbol{\omega},\mathbf{J}\rangle}\right|^2,$$

whose first proposal can be found in Stein (1995) and its multivariate generalization is in Lim and Stein (2008). In Chapter 4, it will be shown that a smoothed fixed domain periodoram with a smoothing kernel $\phi(\cdot)$, written as $\int_{\mathbb{T}^d}\phi(\boldsymbol{\omega})\mathcal{I}^{\mathcal{T}}_{N,\delta}(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega}$, has a nondegenerated limit after appropriate scaling under a mild condition on the spectral density.

## 3.2 Spectral Tail Behavior Model

Recall that we assumed (1.1) for the tail model of the spectral density, which we restate here

$$f(\boldsymbol{\omega}) \sim c|\boldsymbol{\omega}|^{-\alpha}h\left(\frac{\boldsymbol{\omega}}{|\boldsymbol{\omega}|};\alpha,\boldsymbol{\beta}\right) \quad \text{as} \quad |\boldsymbol{\omega}|\to\infty,$$

where $c>0, \alpha>d$ and $h(\mathbf{x};\alpha,\boldsymbol{\beta})$, $\mathbf{x}\in\mathbb{S}^{d-1}$ is a bounded parametrized function defined on a unit sphere in $\mathbb{R}^d$. The function $h$, for instance, enables to incorporate geometric anisotropy into the model. Under this model, we have

$$\begin{aligned}\delta^{d-\alpha}\bar{f}^{\mathcal{T}}_{\delta}(\boldsymbol{\omega}) &\to c\left\{\sum_{j=1}^{d}4\sin^2\left(\frac{\omega_j}{2}\right)\right\}^{2\tau}\sum_{\mathbf{Q}\in\mathbb{Z}^d}|\boldsymbol{\omega}+2\pi\mathbf{Q}|^{-\alpha}h\left(\frac{\boldsymbol{\omega}+2\pi\mathbf{Q}}{|\boldsymbol{\omega}+2\pi\mathbf{Q}|};\alpha,\boldsymbol{\beta}\right)\\ &=: g(\boldsymbol{\omega};c,\alpha,\boldsymbol{\beta}) \qquad (3.1)\end{aligned}$$

for each $\boldsymbol{\omega}\neq\mathbf{0}$ as $\delta\downarrow 0$. Since $(c,\alpha,\boldsymbol{\beta})$ is the set of parameters that affects tail behavior of the spectral density, we call it spectral tail parameters or shortly tail

parameters. The set of the tail parameters will be expressed as $\boldsymbol{\theta} = (c, \alpha, \boldsymbol{\beta}^T)^T$. It can be shown that $g(\boldsymbol{\omega}; \boldsymbol{\theta}) \to 0$ as $\boldsymbol{\omega} \to \mathbf{0}$ when $4\tau - \alpha > 0$. Since we will assume $4\tau - \alpha > 0$, we define $g(\mathbf{0}; \boldsymbol{\theta}) = 0$ hereafter. We often write $g(\cdot; \boldsymbol{\theta})$ or more simply $g(\boldsymbol{\theta})$, omitting its argument $\boldsymbol{\omega}$, to indicate the function $\boldsymbol{\omega} \mapsto g(\boldsymbol{\omega}; \boldsymbol{\theta})$.

## 3.3 Notions and Notations

We finish this chapter with introducing some notions and notations. The $d$-dimensional Fejér kernel with an order $M \in \mathbb{N}$, denoted as $K_M$ is given as

$$\begin{aligned} K_M(\boldsymbol{\omega}) &= \prod_{j=1}^{d} \frac{1}{2\pi M} \frac{\sin^2\left(\frac{M\omega_j}{2}\right)}{\sin^2\left(\frac{\omega_j}{2}\right)} \\ &= \frac{1}{(2\pi M)^d} \left| \sum_{\mathbf{J} \in \mathbb{Z}_M^d} e^{i\langle \boldsymbol{\omega}, \mathbf{J} \rangle} \right|^2 = \frac{1}{(2\pi)^d} \sum_{\mathbf{J}: \|\mathbf{J}\|_\infty \le M} a_{M-|\mathbf{J}|} e^{i\langle \boldsymbol{\omega}, \mathbf{J} \rangle}, \end{aligned}$$

where $\|\cdot\|_\infty$ is a max-norm and $a_{M-|\mathbf{J}|} = \prod_{j=1}^{d} \left(1 - \frac{|J_j|}{M}\right)$. Convolution between two functions is expressed using $*$ so that the convolution of $g$ and $K_M$ is denoted as $g * K_M$. We also use the subscripted expression $g_M$ for simplicity. That is,

$$g_M(\boldsymbol{\omega}) = (g * K_M)(\boldsymbol{\omega}) = \int_{\mathbb{T}^d} g(\boldsymbol{\omega} - \boldsymbol{\upsilon}) K_M(\boldsymbol{\upsilon}) \mathbf{d}\boldsymbol{\upsilon} = \int_{\mathbb{T}^d} g(\boldsymbol{\upsilon}) K_M(\boldsymbol{\omega} - \boldsymbol{\upsilon}) \mathbf{d}\boldsymbol{\upsilon}.$$

When this $g_M$ is again convoluted with another Fejér kernel with an order $L$, this is denoted as $g_{L,M}$. Since the convolution operation is commutative and associative, we have $g_{L,M} = g_M * K_L = g_L * K_M$.

The notation $C^1(\mathbb{T}^d)$ is the set of continuously differentiable function on $\mathbb{T}^d$. When a vector-valued function is in $C^1(\mathbb{T}^d)$, it means each of component functions is in $C^1(\mathbb{T}^d)$. Notation $a(\boldsymbol{\omega}) \lesssim b(\boldsymbol{\omega})$ means that there exists $C > 0$

such that $|a(\boldsymbol{\omega})| \leq C \cdot b(\boldsymbol{\omega})$ for all $\boldsymbol{\omega}$, where $C$ is independent of any variables, functions and arguments on both sides including $a$, $b$ and $\boldsymbol{\omega}$. When $a(\boldsymbol{\omega}) \lesssim b(\boldsymbol{\omega})$ and $b(\boldsymbol{\omega}) \lesssim a(\boldsymbol{\omega})$, we write $a(\boldsymbol{\omega}) \asymp b(\boldsymbol{\omega})$.

# Chapter 4

# Fixed Domain Smoothed Periodogram

## 4.1 Fixed Domain Asymptotics for Smoothed Periodogram

Before we introduce asymptotic properties of the proposed smoothed periodograms, we provide our motivation to use smoothed periodograms in estimating the tail parameter $\boldsymbol{\theta}$.

It can be easily checked that the expected value of periodogram $\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})$ is expressed as $\mathbb{E}\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega}) = (\bar{f}^{\tau}_{\delta} * K_N)(\boldsymbol{\omega})$. If the convergence of $\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega})$ to

$g(\boldsymbol{\omega};\boldsymbol{\theta})$ is uniform in $\boldsymbol{\omega}$, we have

$$\begin{aligned}
&\sup_{\boldsymbol{\omega}\in\mathbb{T}^d}\left|\delta^{d-\alpha}\mathbb{E}\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})-g(\boldsymbol{\omega};\boldsymbol{\theta})\right| \\
&\leq \sup_{\boldsymbol{\omega}\in\mathbb{T}^d}\left|((\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}-g(\boldsymbol{\theta}))*K_N)(\boldsymbol{\omega})\right|+\sup_{\boldsymbol{\omega}\in\mathbb{T}^d}|(g(\boldsymbol{\theta})*K_N)(\boldsymbol{\omega})-g(\boldsymbol{\omega};\boldsymbol{\theta})| \\
&\to 0
\end{aligned}$$

as $N\to\infty$. The second term goes to zero due to the well-known property of the Fejér kernel. See for instance the proof of Proposition 1.15 in Muscalu and Schlag (2013). We can then mimic Whittle's (negative) likelihood approximation for estimating $\boldsymbol{\theta}$ which results in

$$\mathcal{L}_N(\boldsymbol{\theta})=\int_{\mathbb{T}^d}\left\{\frac{\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})}{\delta^{\alpha-d}g(\boldsymbol{\omega};\boldsymbol{\theta})}+\log\left(\delta^{\alpha-d}g(\boldsymbol{\omega};\boldsymbol{\theta})\right)\right\}\mathbf{d}\boldsymbol{\omega}.$$

When the true tail parameter is $\boldsymbol{\theta}_0=(c_0,\alpha_0,\boldsymbol{\beta}_0^T)^T$, the previous argument says that $\mathbb{E}\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})\approx\delta^{\alpha_0-d}g(\boldsymbol{\omega};\boldsymbol{\theta}_0)$ under the true measure. Hence we may expect $\mathcal{L}_N(\boldsymbol{\theta})$ to be close to

$$\mathcal{L}_0(\boldsymbol{\theta})=\int_{\mathbb{T}^d}\left\{\frac{\delta^{\alpha_0-d}g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\delta^{\alpha-d}g(\boldsymbol{\omega};\boldsymbol{\theta})}+\log\left(\delta^{\alpha-d}g(\boldsymbol{\omega};\boldsymbol{\theta})\right)\right\}\mathbf{d}\boldsymbol{\omega}$$

under the true measure. If this is the case, we expect that the minimizer of $\mathcal{L}_N$ would be close to the true tail parameter $\boldsymbol{\theta}_0$. This can be seen from the fact that $\mathcal{L}_0$ is equivalent to

$$\int_{\mathbb{T}^d}\left\{\frac{\delta^{\alpha_0-d}g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\delta^{\alpha-d}g(\boldsymbol{\omega};\boldsymbol{\theta})}-1-\log\left(\frac{\delta^{\alpha_0-d}g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\delta^{\alpha-d}g(\boldsymbol{\omega};\boldsymbol{\theta})}\right)\right\}\mathbf{d}\boldsymbol{\omega}$$

as an objective function for minimization with respect to $\boldsymbol{\theta}$, and the function $x\mapsto x-1-\log x$ is nonnegative and uniquely minimized at $x=1$. These two facts imply that $\boldsymbol{\theta}_0$ minimizes $\mathcal{L}_0(\boldsymbol{\theta})$, which now justifies our motivation for

constructing $\mathcal{L}_N$. If we can prove that the minimizer of $\mathcal{L}_N$, say $\hat{\boldsymbol{\theta}}$, is consistent to $\boldsymbol{\theta}_0$, then we can further obtain an asymptotic order and distribution of the estimator through the classical approximation

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \approx -\left(\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\mathcal{L}_N(\boldsymbol{\theta}_0)\right)^{-1}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\mathcal{L}_N(\boldsymbol{\theta}_0)\right).$$

Answering whether the above approach is theoretically valid eventually boils down to the following problem: for which function $\phi(\boldsymbol{\omega})$ does $\delta^{d-\alpha}\int\phi(\boldsymbol{\omega})\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega}$ becomes close to $\int\phi(\boldsymbol{\omega})g(\boldsymbol{\omega};\boldsymbol{\theta})\mathbf{d}\boldsymbol{\omega}$? Moreover, can we derive the joint limiting distribution of $\int\boldsymbol{\Phi}(\boldsymbol{\omega})\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega} = (\int\phi_1(\boldsymbol{\omega})\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega},\cdots,\int\phi_p(\boldsymbol{\omega})\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega})^T$ for a given $p$-dimensional vector-valued function $\boldsymbol{\Phi} = (\phi_1,\cdots,\phi_p)^T$? We answer these questions with the next theorem, which can be said as the smoothed periodogram theorem under the fixed domain asymptotics. Before we proceed, we would like to mention that this can be seen as a fixed domain asymptotic analogue of Heyde and Gay (1993).

**Theorem 4.1.** *Suppose we have a mean zero stationary Gaussian random field $Z$ on $\mathbb{R}^d$ whose spectral density is $f$ with tail behavior* (1.1)*. Apply the discrete Laplacian operator $\tau$ times to get $Y^{\tau}_{\delta}$ whose spectral density is $\bar{f}^{\tau}_{\delta}$ as defined in Chapter 3. Assume that $\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega})$ and its first partial derivatives with respect to $\boldsymbol{\omega}$ converge uniformly. We also assume that $\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega})$ and its limit $g(\boldsymbol{\omega};\boldsymbol{\theta})$ is in $C^1(\mathbb{T}^d)$. Suppose there is a vector-valued function $\boldsymbol{\Phi} = (\phi_k)_{k=1}^p$ in $C^1(\mathbb{T}^d)$. With the periodogram $\mathcal{I}^{\tau}_{N,\delta}$ defined as in Chapter 3, a fixed domain smoothed periodogram has the following limiting distribution:*

$$N^{d/2}\delta^{d-\alpha}\int_{\mathbb{T}^d}\boldsymbol{\Phi}(\boldsymbol{\omega})\left(\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega}) - \mathbb{E}\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})\right)\mathbf{d}\boldsymbol{\omega} \xrightarrow[N\to\infty]{d} N\left(\mathbf{0},(2\pi)^d\boldsymbol{\Sigma}\right)$$

*where*

$$\boldsymbol{\Sigma} = \int_{\mathbb{T}^d} \left\{ \boldsymbol{\Phi}(\boldsymbol{\omega}) \boldsymbol{\Phi}^T(\boldsymbol{\omega}) + \boldsymbol{\Phi}(\boldsymbol{\omega}) \boldsymbol{\Phi}^T(-\boldsymbol{\omega}) \right\} g^2(\boldsymbol{\omega}; \boldsymbol{\theta}) \mathbf{d}\boldsymbol{\omega}.$$

*The expectation of the fixed domain smoothed periodogram also has the following convergence result:*

$$\delta^{d-\alpha} \int_{\mathbb{T}^d} \boldsymbol{\Phi}(\boldsymbol{\omega}) \mathbb{E} \mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega}) \mathbf{d}\boldsymbol{\omega} \xrightarrow[N \to \infty]{} \int_{\mathbb{T}^d} \boldsymbol{\Phi}(\boldsymbol{\omega}) g(\boldsymbol{\omega}; \boldsymbol{\theta}) \mathbf{d}\boldsymbol{\omega}.$$

The proof relegated in the appendix is based on calculations of the limits of all the joint cumulants. Note that $\phi_1, \cdots, \phi_p \in C^1(\mathbb{T}^d)$ is required which is a stronger condition than integrability or continuity. Intuitively speaking, this is because the integration with a function that resembles a Dirichlet kernel appears in the cumulant representation. It is well known that unlike the Fejér kernel, convolution of a function with a Dirichlet kernel may not converge to the true function value when the function is just integrable or continuous. We can anticipate similar phenomenon from the cumulants of smoothed periodograms, which justifies our possibly restrictive $C^1$ assumption on $\phi_1, \cdots, \phi_p$.

This $C^1$ requirement turns out to be critical indeed for answering whether the usage of $\mathcal{L}_N$ is valid or not. When we use Theorem 4.1 to answer this question, the reciprocal of $g(\cdot; \boldsymbol{\theta})$ is used as $\phi$. However, we have $g(\boldsymbol{\omega}; \boldsymbol{\theta}) \asymp |\boldsymbol{\omega}|^{4\tau - \alpha}$ under the tail assumption (1.1). Then the reciprocal of $g(\boldsymbol{\omega}; \boldsymbol{\theta})$ explodes near the origin under the assumption $4\tau - \alpha > 0$, from which $\frac{1}{g(\cdot; \boldsymbol{\theta})} \notin C^1(\mathbb{T}^d)$. Hence we cannot apply Theorem 4.1 to verify whether $\int \frac{\delta^{d-\alpha_0} \mathcal{I}^{\tau}_{N,\delta}}{g(\cdot; \boldsymbol{\theta})}$ behaves like $\int \frac{g(\cdot; \boldsymbol{\theta}_0)}{g(\cdot; \boldsymbol{\theta})}$ when the true tail parameter is $\boldsymbol{\theta}_0$.

## 4.2 Tapered Periodogram and Smoothed Periodogram with a compactly supported kernel

Although the previous attempt to estimate $\boldsymbol{\theta}$ using $\mathcal{L}_N$ is not successful, it helps us to devise a suitable kernel to construct a Whittle type objective function. Suppose we choose $\phi(\cdot) = \mathcal{K}(\cdot - \boldsymbol{\omega})$ for some $\boldsymbol{\omega} \in \mathbb{T}^d$, where $\mathcal{K}$ is a $C^1$ function on $\mathbb{T}^d$ whose mass is mostly concentrated around the origin. Then, Theorem 4.1 provides theoretical background that $\delta^{d-\alpha} \int \mathcal{K}(\mathbf{x} - \boldsymbol{\omega}) \mathcal{I}^{\tau}_{N,\delta}(\mathbf{x}) \mathbf{dx}$ will approach to $\int \mathcal{K}(\mathbf{x} - \boldsymbol{\omega}) g(\mathbf{x}; \boldsymbol{\theta}) \mathbf{dx}$ as $N \to \infty$. We can expect that $\int \mathcal{K}(\mathbf{x} - \boldsymbol{\omega}) g(\mathbf{x}; \boldsymbol{\theta}) \mathbf{dx}$ would contain essential information about $g(\boldsymbol{\omega}; \boldsymbol{\theta})$ since the mass of $\mathcal{K}(\cdot - \boldsymbol{\omega})$ is mostly concentrated around $\boldsymbol{\omega}$. Denote $\int \mathcal{K}(\mathbf{x} - \boldsymbol{\omega}) \mathcal{I}^{\tau}_{N,\delta}(\mathbf{x}) \mathbf{dx}$ as $\hat{\mathcal{I}}^{\tau}_{N,\delta,\mathcal{K}}(\boldsymbol{\omega})$ and $\int \mathcal{K}(\mathbf{x} - \boldsymbol{\omega}) g(\mathbf{x}; \boldsymbol{\theta}) \mathbf{dx}$ as $g_{\mathcal{K}}(\boldsymbol{\omega}; \boldsymbol{\theta})$. By collecting $\hat{\mathcal{I}}^{\tau}_{N,\delta,\mathcal{K}}(\boldsymbol{\omega})$ from many distinct $\boldsymbol{\omega} \in \Omega \subset \mathbb{T}^d$, we can gather the information about $g(\boldsymbol{\omega}; \boldsymbol{\theta})$ at $\boldsymbol{\omega} \in \Omega$ in the form of $g_{\mathcal{K}}(\boldsymbol{\omega}; \boldsymbol{\theta})$. This can be exploited to estimate the tail parameter. Assuming that the true tail parameter is $\boldsymbol{\theta}_0$, Theorem 4.1 implies that $\delta^{d-\alpha_0} \hat{\mathcal{I}}^{\tau}_{N,\delta,\mathcal{K}}(\boldsymbol{\omega})$ gets close to $g_{\mathcal{K}}(\boldsymbol{\omega}; \boldsymbol{\theta}_0)$ as $N \to \infty$. Then by seeking $\hat{\boldsymbol{\theta}}$ that makes $\hat{\mathcal{I}}^{\tau}_{N,\delta,\mathcal{K}}(\boldsymbol{\omega})$ and $\delta^{\hat{\alpha}-d} g_{\mathcal{K}}(\boldsymbol{\omega}; \hat{\boldsymbol{\theta}})$ as close as possible for all $\boldsymbol{\omega} \in \Omega$, we could have $\hat{\boldsymbol{\theta}}$ close to $\boldsymbol{\theta}_0$. In fact, this approach is the essence of the method we propose, which will be theoretically investigated in Chapter 5.

There are several ways to choose a kernel $\mathcal{K}$. One is to use a compactly supported function. For instance, assume that there is a function $K \in C^1(\mathbb{T}^d)$ such that $K \geq 0, K(\mathbf{x}) = K(-\mathbf{x}), \int K(\mathbf{x}) \mathbf{dx} = 1$ and $K(\mathbf{x}) = 0$ whenever $\|\mathbf{x}\|_\infty \geq 1$. With $\mathcal{K}(\cdot) = \frac{1}{h^d} K\left(\frac{\cdot}{h}\right)$ for small $h > 0$, we can get a suitable kernel function. In this case, we write $\hat{\mathcal{I}}^{\tau}_{N,\delta,\mathcal{K}}(\boldsymbol{\omega})$ as $\hat{\mathcal{I}}^{\tau}_{N,\delta,h}(\boldsymbol{\omega})$ to show the dependency of the

smoothed periodogram on the bandwidth $h$ clear, and call it the *smoothed periodogram with the kernel $K$ and the bandwidth $h$*. The benefit of this construction is that the corresponding smoothed periodograms become asymptotically uncorrelated. Observe that Theorem 4.1 implies that

$$\begin{aligned}&\operatorname{cov}\left(\delta^{d-\alpha}\hat{\mathcal{I}}^{\tau}_{N,\delta,h}(\boldsymbol{\omega}), \delta^{d-\alpha}\hat{\mathcal{I}}^{\tau}_{N,\delta,h}(\boldsymbol{\upsilon})\right)\\ &\to \left(\frac{2\pi}{h^2}\right)^d \int_{\mathbb{T}^d} \left\{K\left(\frac{\mathbf{x}-\boldsymbol{\omega}}{h}\right)K\left(\frac{\mathbf{x}-\boldsymbol{\upsilon}}{h}\right) + K\left(\frac{\mathbf{x}-\boldsymbol{\omega}}{h}\right)K\left(\frac{-\mathbf{x}-\boldsymbol{\upsilon}}{h}\right)\right\} g^2(\mathbf{x};\boldsymbol{\theta})\mathbf{dx}\end{aligned}$$

as $N \to \infty$. Because of the properties of $K$, the right hand side becomes 0 if $\|\boldsymbol{\omega} \pm \boldsymbol{\upsilon}\|_\infty \geq 2h$. By taking $\Omega \subset \mathbb{T}^d \cap \{\boldsymbol{\omega} : 0 < \omega_1 < \pi\}$ to satisfy $\|\boldsymbol{\omega} - \boldsymbol{\upsilon}\|_\infty \geq 2h$ for all $\boldsymbol{\omega}, \boldsymbol{\upsilon} \in \Omega$ such that $\boldsymbol{\omega} \neq \boldsymbol{\upsilon}$, the smoothed periodograms $\{\hat{\mathcal{I}}_{N,\delta,h}(\boldsymbol{\omega}) : \boldsymbol{\omega} \in \Omega\}$ are asymptotically mutually independent, which is a handy property to work with.

Another is to use the Fejér kernel. Its advantage is revealed in the evaluation of $\hat{\mathcal{I}}^{\tau}_{N,\delta,\mathcal{K}}(\boldsymbol{\omega})$ in practice. From the definition of $\hat{\mathcal{I}}^{\tau}_{N,\delta,\mathcal{K}}$, we can see that

$$\begin{aligned}\hat{\mathcal{I}}^{\tau}_{N,\delta,\mathcal{K}}(\boldsymbol{\omega}) &= \frac{1}{(2\pi N)^d}\int_{\mathbb{T}^d} \mathcal{K}(\mathbf{x}-\boldsymbol{\omega})\left|\sum_{\mathbf{J}\in\mathbb{Z}^d_N} Y^{\tau}_{\delta}(\mathbf{J})e^{-i\langle \mathbf{x},\mathbf{J}\rangle}\right|^2 \mathbf{dx}\\ &= \frac{1}{(2\pi)^d}\sum_{\mathbf{J}:\|\mathbf{J}\|_\infty \leq N}\left\{\frac{1}{N^d}\sum_{\mathbf{K}:\mathbf{J}+\mathbf{K},\mathbf{K}\in\mathbb{Z}^d_N} Y^{\tau}_{\delta}(\mathbf{J}+\mathbf{K})Y^{\tau}_{\delta}(\mathbf{K})\right\}\int_{\mathbb{T}^d}\mathcal{K}(\mathbf{x}-\boldsymbol{\omega})e^{-i\langle \mathbf{x},\mathbf{J}\rangle}\mathbf{dx}\\ &=: \frac{1}{(2\pi)^d}\sum_{\mathbf{J}:\|\mathbf{J}\|_\infty\leq N}\hat{C}^{\tau}_{\delta}(\mathbf{J})\hat{\mathcal{K}}(\mathbf{J})e^{-i\langle\boldsymbol{\omega},\mathbf{J}\rangle},\end{aligned}$$

where $\hat{\mathcal{K}}$ is the Fourier coefficient of $\mathcal{K}$ given as

$$\hat{\mathcal{K}}(\mathbf{J}) = \int_{\mathbb{T}^d}\mathcal{K}(\mathbf{x})e^{-i\langle\mathbf{x},\mathbf{J}\rangle}\mathbf{dx}.$$

Now, consider to take $\mathcal{K} = K_M$, where $K_M$ is the Fejér kernel of an order $M$, for some large enough $M \in \mathbb{N}$ yet $M$ much less than $N$. We can write

$\hat{\mathcal{I}}^{\tau}_{N,\delta,\mathcal{K}} = \hat{\mathcal{I}}^{\tau}_{N,\delta,M}$ to explicitly show its dependency to $M$. Since

$$K_M(\mathbf{x}) = \frac{1}{(2\pi M)^d}\left|\sum_{\mathbf{J}\in\mathbb{Z}^d_M} e^{i\langle \mathbf{x},\mathbf{J}\rangle}\right|^2 = \frac{1}{(2\pi)^d}\sum_{\mathbf{J}:\|\mathbf{J}\|_\infty\le M} a_{M-|\mathbf{J}|}e^{i\langle \mathbf{x},\mathbf{J}\rangle}$$

where $a_{M-|\mathbf{J}|} = \prod_{j=1}^d \left(1-\frac{|J_j|}{M}\right)$, we have

$$\hat{K}_M(\mathbf{J}) = \begin{cases} a_{M-|\mathbf{J}|}, & \|\mathbf{J}\|_\infty \le M, \\ 0, & \text{otherwise.} \end{cases}$$

Then $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ becomes

$$\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d}\sum_{\mathbf{J}:\|\mathbf{J}\|_\infty\le M} a_{M-|\mathbf{J}|}\hat{C}^{\tau}_{\delta}(\mathbf{J})e^{-i\langle\boldsymbol{\omega},\mathbf{J}\rangle}.$$

The benefit of this choice is that it is enough to store $\hat{C}^{\tau}_{\delta}(\mathbf{J})$ for $\|\mathbf{J}\|_\infty \le M$ to evaluate $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ for any $\boldsymbol{\omega}\in\mathbb{T}^d$. There are only $(2M-1)^d$ values of such $\hat{C}^{\tau}_{\delta}(\mathbf{J})$, which is much less than $N^d$ so that it saves our memory in computation in practice. Furthermore, if we set $\Omega = \left\{\frac{2\pi\mathbf{K}}{M} : \mathbf{K}\in\mathbb{Z}^d\right\}$ and evaluate $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ at $\boldsymbol{\omega}\in\Omega$, we can exploit the fast Fourier transform algorithm.

Compared to the original periodogram $\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})$ which can be expressed as

$$\begin{aligned}\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega}) &= \frac{1}{(2\pi N)^d}\left|\sum_{\mathbf{J}\in\mathbb{Z}^d_N} Y^{\tau}_{\delta}(\mathbf{J})e^{-i\langle\boldsymbol{\omega},\mathbf{J}\rangle}\right|^2 \\ &= \frac{1}{(2\pi)^d}\sum_{\|\mathbf{J}\|_\infty\le N}\left\{\frac{1}{N^d}\sum_{\mathbf{K}:\mathbf{J},\mathbf{J}+\mathbf{K}\in\mathbb{Z}^d_N} Y^{\tau}_{\delta}(\mathbf{J}+\mathbf{K})Y^{\tau}_{\delta}(\mathbf{K})\right\}e^{-i\langle\boldsymbol{\omega},\mathbf{J}\rangle} \\ &=: \frac{1}{(2\pi)^d}\sum_{\|\mathbf{J}\|_\infty\le N}\hat{C}^{\tau}_{\delta}(\mathbf{J})e^{-i\langle\boldsymbol{\omega},\mathbf{J}\rangle},\end{aligned}$$

the summation in $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ is up to $M$ which can be much less than $N$. Hence, we call $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ as the *tapered periodogram* of an order $M$. One property of

the tapered periodogram is that it is guaranteed to be nonnegative. Although $\mathcal{I}^{\tau}_{N,\delta}$ is always nonnegative, there is no positivity restriction on $\phi$ in Theorem 4.1 so that the corresponding smoothed periodogram can be negative for general $\phi$. However, the Fejér kernel, which is always nonnegative, guarantees the tapered periodogram to be nonnegative. The smoothed periodogram with a compactly supported kernel also has this property since we use a kernel $K$ satisfying $K \geq 0$.

Closing the current chapter, we would like to leave two comments. The first is that the order $M$ for the tapered periodogram and the bandwidth $h$ for the smoothed periodogram with a compactly supported kernel are considered to be fixed throughout this work. We may expect $\delta^{d-\alpha}\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ and $\delta^{d-\alpha}\hat{\mathcal{I}}^{\tau}_{N,\delta,h}(\boldsymbol{\omega})$ to converge toward $g(\boldsymbol{\omega};\boldsymbol{\theta})$ for each $\boldsymbol{\omega}$ by letting $M$ increase to infinity or $h$ decrease to zero along $N$, yet this needs further theoretical investigation to be justified. The other is that under the fixed domain asymptotics, the smoothed periodogram $\int \phi(\boldsymbol{\omega})\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega}$ needs to be scaled by $\delta^{d-\alpha}$ to have a nondegenerate limit $\int \phi(\boldsymbol{\omega})g(\boldsymbol{\omega};\boldsymbol{\theta})\mathbf{d}\boldsymbol{\omega}$ which is implied by Theorem 4.1. When we have little knowledge for $\alpha$, the scaling factor becomes an unknown quantity. Hence the smoothed periodogram itself cannot be used as a nonparametric estimator in this case. However, we can estimate $\alpha$ consistently under the tail assumption (1.1) and some regularity assumptions, which is proved in Chapter 5.

# Chapter 5

# Tail Parameter Estimation

## 5.1 Estimation and Fixed Domain Asymptotic Result

Based on the discussion in Chapter 4, we compare $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ or $\hat{\mathcal{I}}^{\tau}_{N,\delta,h}(\boldsymbol{\omega})$ with $\delta^{\alpha-d}g_M(\boldsymbol{\omega};\boldsymbol{\theta})$ to estimate $\boldsymbol{\theta}$, where $g_M(\cdot;\boldsymbol{\theta}) = g(\cdot;\boldsymbol{\theta})*K_M$. Consider to minimize

$$\sum_{\boldsymbol{\omega}\in\Omega_M}\left\{\frac{\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})}{\delta^{\alpha-d}g_M(\boldsymbol{\omega};\boldsymbol{\theta})}+\log\left(\delta^{\alpha-d}g_M(\boldsymbol{\omega};\boldsymbol{\theta})\right)\right\}$$

with respect to $\boldsymbol{\theta}$ for some prespecified finite set $\Omega_M$. The idea for this construction is similar to $\mathcal{L}_N$ introduced in the beginning of Chapter 4. Speaking the conclusion first, however, the minimizer of the above has an $N^{-1}$-order bias. This bias is the same type of bias with the one introduced in Guyon (1982) and is called the edge effect bias. The edge effect bias in the fixed domain asymptotics is originated from the convergence rate of $\delta^{d-\alpha}\mathbb{E}\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ toward

$g_M(\boldsymbol{\omega};\boldsymbol{\theta})$. The brief description is that we have

$$\mathbb{E}\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega}) = \bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega}) := (\bar{f}^{\tau}_{\delta} * K_N * K_M)(\boldsymbol{\omega})$$

which is getting close to $g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}) = (g(\boldsymbol{\theta}) * K_N * K_M)(\boldsymbol{\omega})$ after scaling by $\delta^{d-\alpha}$ with a speed of $\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega})$ getting close to $g(\boldsymbol{\omega};\boldsymbol{\theta})$, while $g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})$ goes to $g_M(\boldsymbol{\omega};\boldsymbol{\theta})$ with a speed of $N^{-1}$. That is to say, even when $\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega})$ goes to $g(\boldsymbol{\omega};\boldsymbol{\theta})$ faster than $N^{-1}$, the difference between $\delta^{d-\alpha}\mathbb{E}\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ and $g_M(\boldsymbol{\omega};\boldsymbol{\theta})$ is of order $N^{-1}$. To alleviate the edge effect, we propose to replace $g_M$ in the above objective function with $g_{N,M}$. The replacement with $g_{N,M}$ just removes the source of $N^{-1}$-order bias so that it enhances the performance of the estimator. Note that when we are working with $\hat{\mathcal{I}}^{\tau}_{N,\delta,h}$, the smoothed periodogram with a kernel $K$ and a bandwidth $h$, $g_{N,h}$ is used instead of $g_{N,M}$ where

$$g_{N,h}(\boldsymbol{\omega};\boldsymbol{\theta}) = \int_{\mathbb{T}^d}\int_{\mathbb{T}^d} g(\boldsymbol{\upsilon})K_N(\mathbf{x}-\boldsymbol{\upsilon})\frac{1}{h^d}K\left(\frac{\mathbf{x}-\boldsymbol{\omega}}{h}\right)\mathbf{d}\boldsymbol{\upsilon}\mathbf{d}\mathbf{x}.$$

When the tapered periodogram is in use, we take $\Omega_M = \left\{\frac{2\pi\mathbf{J}}{M} : \mathbf{J} \in \mathbb{Z}^d_M\right\} \setminus (-t,t)^d \subset \mathbb{T}^d$ for some given small $t > 0$. Benefit of choosing frequencies of the form $\frac{2\pi\mathbf{J}}{M}$ has been already discussed in the previous section. Puncturing some small set around the origin is to avoid values close to the origin in the denominator. Note that $g(\boldsymbol{\omega};\boldsymbol{\theta})$ is defined as zero and continuous at the origin when $4\tau - \alpha > 0$, which makes $g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})$ also close to zero around the origin as well when both $N$ and $M$ are large. When the smoothed periodogram with a compactly supported kernel is in use, we take $h = \frac{\pi}{M}$ for some large $M \in \mathbb{N}$ and set $\Omega_h = \Omega_M = \left\{2h\mathbf{J} : \mathbf{J} \in \mathbb{Z}^d_M, 0 < J_1 < \frac{M}{2}\right\} \setminus (-t,t)^d$. This choice not only allows us to exploit fast Fourier transform for evaluating smoothed periodograms but

also makes the corresponding smoothed periodograms asymptotically uncorrelated. By the choice of $h$ depending on the natural number $M$, we may write $\hat{\mathcal{I}}^{\tau}_{N,\delta,h}$ as $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}$, which is the same notation for tapered periodogram. From now and below, we will use $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}, g_{N,M}$ and $\Omega_M$ to indicate both versions of objective functions.

Along with the above discussion, we arrive at the next objective function which can be interpreted as a locally smoothed version of negative Whittle likelihood:

$$L_N(\boldsymbol{\theta}) = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \left\{ \frac{\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})}{\delta^{\alpha-d} g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})} + \log\left(\delta^{\alpha-d} g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})\right) \right\}.$$

In order to prove that the minimizer of $L_N$ consistently estimates the tail parameter, we impose mild regularity assumptions on the parameter space and the limiting function $g(\boldsymbol{\omega};\boldsymbol{\theta})$ in the following:

**Assumption 1.**

*(A1) The parameter space is* $\Theta = \mathbb{R}^+ \times \mathcal{A} \times \mathcal{B}$ *so that* $c \in \mathbb{R}^+, \alpha \in \mathcal{A}$ *and* $\boldsymbol{\beta} \in \mathcal{B}$. $\mathcal{A}$ *is a compact interval* $[\alpha_L, \alpha_U] \subset (d, 4\tau)$ *and* $\mathcal{B}$ *is a compact set with a nonempty interior* $\mathcal{B}^\circ$. *We also assume that* $\{g(\boldsymbol{\omega};\boldsymbol{\theta}) : \boldsymbol{\omega} \in \mathbb{T}^d\}$ *is identifiable on* $\Theta$, *i.e., if there exists* $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ *satisfying* $g(\boldsymbol{\omega};\boldsymbol{\theta}) = g(\boldsymbol{\omega};\boldsymbol{\theta}')$ *almost everywhere on* $\boldsymbol{\omega} \in \mathbb{T}^d$, *we have* $\boldsymbol{\theta} = \boldsymbol{\theta}'$. *Recall that* $g(\boldsymbol{\omega};\boldsymbol{\theta})$ *is of the form*

$$g(\boldsymbol{\omega};\boldsymbol{\theta}) = c\left\{\sum_{j=1}^{d} 4\sin^2\left(\frac{\omega_j}{2}\right)\right\}^{2\tau} \sum_{\mathbf{Q}\in\mathbb{Z}^d} |\boldsymbol{\omega} + 2\pi\mathbf{Q}|^{-\alpha} h\left(\frac{\boldsymbol{\omega}}{|\boldsymbol{\omega}|}; \alpha, \boldsymbol{\beta}\right)$$

*with* $\boldsymbol{\theta} = (c, \alpha, \boldsymbol{\beta}^T)^T$ *when* $\boldsymbol{\omega} \neq \mathbf{0}$ *and* $g(\mathbf{0};\boldsymbol{\theta})$ *is defined as zero.*

*(A2) There exists a constant* $\gamma > 0$ *such that*

$$\sup_{\boldsymbol{\omega}\in\mathbb{T}^d} \left| \delta^{d-\alpha} \bar{f}_\delta^{\mathcal{T}}(\boldsymbol{\omega}) - g(\boldsymbol{\omega};\boldsymbol{\theta}) \right| = O(\delta^\gamma)$$

*for each* $\boldsymbol{\theta} \in \Theta$*. We allow* $\gamma = \gamma(\boldsymbol{\theta})$ *to depend on* $\boldsymbol{\theta}$*. Moreover, all the first partial derivatives of* $\delta^{d-\alpha}(\boldsymbol{\omega})$ *with respect to* $\boldsymbol{\omega}$ *converge uniformly.*

*(A3)* $g(\cdot; 1, \alpha, \boldsymbol{\beta}) \in C^1(\mathbb{T}^d)$*,*

$$\sup \left\{ g(\boldsymbol{\omega}; 1, \alpha, \boldsymbol{\beta}) : \boldsymbol{\omega} \in \mathbb{T}^d, (\alpha, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{B} \right\} < \infty$$

*and*

$$\sup \left\{ \left| \frac{\partial g}{\partial \omega_j}(\boldsymbol{\omega}; 1, \alpha, \boldsymbol{\beta}) \right| : \boldsymbol{\omega} \in \mathbb{T}^d, (\alpha, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{B}, j = 1, \cdots, d \right\} < \infty.$$

*(A4) For each* $t > 0$*,*

$$\inf \left\{ g(\boldsymbol{\omega}; 1, \alpha, \boldsymbol{\beta}) : \boldsymbol{\omega} \in \mathbb{T}^d \setminus (-t, t)^d, (\alpha, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{B} \right\} > 0.$$

*(A5)* $(\alpha, \boldsymbol{\beta}) \mapsto g(\boldsymbol{\omega}; 1, \alpha, \boldsymbol{\beta})$ *is continuous on* $\mathcal{A} \times \mathcal{B}$ *for each* $\boldsymbol{\omega} \in \mathbb{T}^d$*.*

*(A6)* $\frac{\partial g(\cdot;1,\alpha,\boldsymbol{\beta})}{\partial(\alpha,\boldsymbol{\beta})}$ *and* $\frac{\partial^2 g(\cdot;1,\alpha,\boldsymbol{\beta})}{\partial(\alpha,\boldsymbol{\beta})\partial(\alpha,\boldsymbol{\beta})^T}$ *are* $C^1(\mathbb{T}^d)$*, continuous with respect to* $(\alpha, \boldsymbol{\beta})$ *for each* $\boldsymbol{\omega} \in \mathbb{T}^d$*, and uniformly bounded on* $\boldsymbol{\omega} \in \mathbb{T}^d$ *and* $(\alpha, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{B}$*.*

Under the above assumptions, we prove that the minimizer of $L_N$ is a consistent estimator for the tail parameter under the fixed domain asymptotics and derive the asymptotic order of the estimator as well. For some cases, we can also derive the asymptotic distribution of the estimator. These results are stated in the following theorem.

**Theorem 5.1.** *Suppose we have a mean zero stationary random field $Z$ on $\mathbb{R}^d$ whose spectral density is $f$ with the tail behavior* (1.1)*. Apply the discrete Laplacian operator $\tau$ times to obtain $Y_\delta^\tau$ whose spectral density is $\bar{f}_\delta^\tau$ which is described in Chapter 3. With the periodogram $\mathcal{I}_{N,\delta}^\tau$ which is defined in Chapter 3, $M \in \mathbb{N}$ and a small positive number $t > 0$, consider $\hat{\mathcal{I}}_{N,\delta,M}^\tau(\boldsymbol{\omega})$, which represents either the tapered periodogram of an order $M$ or the smoothed periodogram with a kernel $K$ and a bandwidth $h$ where both are introduced in Chapter 4. Here, a kernel $K$ should satisfy $K \in C^1(\mathbb{T}^d), K \geq 0, K(\mathbf{x}) = K(-\mathbf{x}), \int K(\mathbf{x})\mathbf{dx} = 1$ and $K(\mathbf{x}) = 0$ whenever $\|\mathbf{x}\|_\infty \geq 1$. Consider $\hat{\mathcal{I}}_{N,\delta,M}^\tau(\boldsymbol{\omega})$ at frequencies $\boldsymbol{\omega} \in \Omega_M$ where $\Omega_M = \left\{\frac{2\pi\mathbf{J}}{M} : \mathbf{J} \in \mathbb{Z}^d\right\} \setminus (-t,t)^d \subset \mathbb{T}^d$ for the tapered periodogram and $\Omega_M = \Omega_h = \left\{2h\mathbf{J} : \mathbf{J} \in \mathbb{Z}_M^d, 0 < J_1 < \frac{M}{2}\right\} \setminus (-t,t)^d$ with $h = \frac{\pi}{M}$ for the smoothed periodogram with a compactly supported kernel. Define*

$$L_N(\boldsymbol{\theta}) = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \left\{ \frac{\hat{\mathcal{I}}_{N,\delta,M}^\tau(\boldsymbol{\omega})}{\delta^{\alpha-d} g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})} + \log\left(\delta^{\alpha-d} g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})\right) \right\}$$

*where $g_{N,M}(\boldsymbol{\theta}) = g(\boldsymbol{\theta}) * K_N * K_M$ for the tapered periodogram case and $g_{N,M}(\boldsymbol{\theta}) = g_{N,h}(\boldsymbol{\theta}) = g(\boldsymbol{\theta}) * K_N * \frac{1}{h^d} K\left(\frac{\cdot}{h}\right)$ for the smoothed periodogram with a compactly supported kernel case. Consider the Assumption 1 and assume that the true tail parameter $\boldsymbol{\theta}_0$ lies in the interior of $\Theta$. Then there exists $M_0 \in \mathbb{N}$ such that for every $M \geq M_0$, the minimizer $\hat{\boldsymbol{\theta}}$ of $L_N(\boldsymbol{\theta})$ converges to $\boldsymbol{\theta}_0$ in probability as $N \to \infty$. Specifically,*

$$N^{\min\{\frac{d}{2},\gamma\}} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} = O_P(1)$$

*and*

$$N^{\min\{\frac{d}{2},\gamma\}} (\log N)^{-1} (\log \hat{c} - \log c_0) = N^{\min\{\frac{d}{2},\gamma\}} (\hat{\alpha} - \alpha_0) + o_P(1).$$

*If* $\frac{d}{2} < \gamma$, *then*

$$N^{d/2}\begin{pmatrix}\hat{\alpha}-\alpha_0\\ \hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\end{pmatrix} \xrightarrow[N\to\infty]{d} N\left(\mathbf{0}, (2\pi)^d[\mathcal{J}^{-1}\Sigma\mathcal{J}^{-1}]_{-1,-1}\right)$$

*where* $[A]_{-1,-1}$ *is the matrix made by removing the first row and the first column each from the matrix* $A$, *and* $\mathcal{J}$ *and* $\Sigma$ *are defined as*

$$\mathcal{J} = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M}\left(\frac{\partial \log g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right)^T$$

*and*

$$\Sigma = \int_{\mathbb{T}^d}\left\{\boldsymbol{\Phi}_M(\mathbf{x})\boldsymbol{\Phi}_M^T(\mathbf{x}) + \boldsymbol{\Phi}_M(\mathbf{x})\boldsymbol{\Phi}_M^T(-\mathbf{x})\right\} g^2(\mathbf{x};\boldsymbol{\theta}_0)\mathbf{dx}$$

*where*

$$\boldsymbol{\Phi}_M(\mathbf{x}) = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M}\frac{\partial \log g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{K_M(\boldsymbol{\omega}-\mathbf{x})}{g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}$$

*when the tapered periodogram is used. When the smoothed periodogram with a compactly supported kernel is used, then the same result holds with*

$$\mathcal{J} = (2h)^d \sum_{\boldsymbol{\omega}\in\Omega_h}\left(\frac{\partial \log g_h(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log g_h(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right)^T$$

*and*

$$\Sigma = (2h)^{2d}\sum_{\boldsymbol{\omega}\in\Omega_h}\left(\frac{\partial \log g_h(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log g_h(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right)^T \frac{1}{g_h^2(\boldsymbol{\omega};\boldsymbol{\theta}_0)} \times \int_{\mathbb{T}^d} g^2(\mathbf{x};\boldsymbol{\theta}_0)\left\{\frac{1}{h^d}K\left(\frac{\boldsymbol{\omega}-\mathbf{x}}{h}\right)\right\}^2\mathbf{dx},$$

*where* $g_h(\boldsymbol{\theta}) = g(\boldsymbol{\theta}) * \frac{1}{h^d}K\left(\frac{\cdot}{h}\right)$. *The limiting distribution of* $N^{d/2}(\log N)^{-1}(\log\hat{c} - \log c_0)$ *is identical to that of* $N^{d/2}(\hat{\alpha}-\alpha_0)$.

*Sketch of the proof.* The detailed proof is deferred to the appendix. We briefly sketch the idea of the proof with some explanations about the Assumption 1.

The proof takes a detour: showing the consistency and asymptotic order of the estimator which is obtained by minimizing $L_{0N}(\boldsymbol{\theta})$, instead of $L_N(\boldsymbol{\theta})$, where $L_{0N}(\boldsymbol{\theta})$ is defined as

$$L_{0N}(\boldsymbol{\theta}) = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \left\{ \frac{\delta^{d-\alpha_0}\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})}{g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})} + \log g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}) \right\}.$$

$\alpha_0$ is the true value of $\alpha$. Be aware that although this requires the knowledge of $\alpha_0$, we still have $\alpha$ as a variable in minimization so that its estimator can be obtained. This seems odd at first glance, but it is needed for extending the result to cover the case for unknown $\alpha$.

Recall that from our tail behavior model (1.1), $\boldsymbol{\theta} = (c, \alpha, \boldsymbol{\beta}^T)^T$. By the profiling approach and the compactness of $\mathcal{A}$ and $\mathcal{B}$ in the assumption (A1), it is shown that $\Theta' \subset \Theta$ can be taken as a compact set so that

$$\mathbb{P}\left( \arg\min_{\boldsymbol{\theta}\in\Theta} L_{0N}(\boldsymbol{\theta}) \in \Theta' \right) \to 1$$

as $N \to \infty$.

Next we show that

$$L_{0N}(\boldsymbol{\theta}) \approx \int_{\bar{\Omega}} \left\{ \frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})} + \log g(\boldsymbol{\omega};\boldsymbol{\theta}) \right\} \mathbf{d}\boldsymbol{\omega} \tag{5.1}$$

where $\bar{\Omega} = \mathbb{T}^d \setminus (-t,t)^d$ for the tapered periodogram case and $\bar{\Omega} = \{\boldsymbol{\omega} \in \mathbb{T}^d : 0 < \omega_1 < \pi\} \setminus (-t,t)^d$ for the smoothed periodogram with a compactly supported kernel case. To show this, the assumption (A2) is used (actually its weaker version is enough. See Proposition B.3 for details) to make use of

Theorem 4.1 so that $\delta^{d-\alpha_0}\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega}) \approx g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)$. ($g_M(\boldsymbol{\theta})$ should be understood as $g_h(\boldsymbol{\theta}) = g(\boldsymbol{\theta}) * \frac{1}{h^d}K\left(\frac{\cdot}{h}\right)$ when the smoothed periodogram with a compactly supported kernel is used.) Next, the assumption (A3) is used to ensure the approximation $g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}) \approx g_M(\boldsymbol{\omega};\boldsymbol{\theta})$ and further $g_M(\boldsymbol{\omega};\boldsymbol{\theta}) \approx g(\boldsymbol{\omega};\boldsymbol{\theta})$. The assumption (A4) controls $g(\boldsymbol{\omega};\boldsymbol{\theta})$ in the denominator.

We now claim that the minimizer of $L_{0N}(\boldsymbol{\theta})$ is asymptotically close enough to the minimizer of the right hand side of (5.1). As an objective function, it is equivalent to

$$\int_{\bar{\Omega}} \left\{ \frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})} - 1 - \log \frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})} \right\} \mathbf{d}\boldsymbol{\omega}.$$

This is minimized when $g(\boldsymbol{\omega};\boldsymbol{\theta}_0) = g(\boldsymbol{\omega};\boldsymbol{\theta})$ for almost every $\boldsymbol{\omega} \in \bar{\Omega}$. The continuity with respect to $\boldsymbol{\theta}$ (the assumption (A5)) and the identifiability (the assumption (A1)) are now applied to complete the proof for consistency of the minimizer of $L_{0N}(\boldsymbol{\theta})$.

The asymptotic order and normality just follow from the classical delta method and the proof is in Proposition B.4. Assumptions (A2) and (A6) are involved in this process.

Now we work with $L_N(\boldsymbol{\theta})$, a version with unknown $\alpha$. It is immediate that $L_N(\boldsymbol{\theta}) = L_N(c,\alpha,\boldsymbol{\beta})$ is equivalent to $L_{0N}(c\delta^{\alpha-\alpha_0},\alpha,\boldsymbol{\beta})$ as an objective function. Hence denoting the minimizer of $L_N(\boldsymbol{\theta})$ as $\hat{\boldsymbol{\theta}} = (\hat{c},\hat{\alpha},\hat{\boldsymbol{\beta}}^T)^T$, $(\hat{c}\delta^{\hat{\alpha}-\alpha_0},\hat{\alpha},\hat{\boldsymbol{\beta}}^T)^T$ is consistent to $(c_0,\alpha_0,\boldsymbol{\beta}_0^T)^T$ and its convergence order is $O_P(N^{-\min\{\frac{d}{2},\gamma\}})$ where $\gamma$ is a constant in the assumption (A2). Especially, we have $\hat{\alpha}-\alpha_0 = O_P(N^{-\min\{\frac{d}{2},\gamma\}})$ from which we have $(\hat{\alpha}-\alpha_0)\log\delta = O_P(N^{-\min\{\frac{d}{2},\gamma\}}\log N)$. Since we also have $\log\hat{c}+(\hat{\alpha}-\alpha_0)\log\delta-\log c_0 = O_P(N^{-\min\{\frac{d}{2},\gamma\}})$, we now have $N^{\min\{\frac{d}{2},\gamma\}}\log N(\log\hat{c}-$

$\log c_0) = \hat{\alpha} - \alpha_0 + O_P((\log N)^{-1})$. □

## 5.2 Examples with Parametric Models

When we only impose the tail behavior condition on the spectral density $f$ and do not have the exact form of it, we cannot check whether the assumption (A2) actually holds or not. However, we can check (A2) for available parametric covariance models. In the next three examples, we will describe some parametric models with their $\gamma$ in (A2). Moreover, we briefly compare our results to the relevant preexisting works.

**Example 5.2.** *Recall that the isotropic Matérn covariance function has the spectral density*

$$f(\boldsymbol{\omega}; \sigma^2, \nu, a) = \frac{\sigma^2 a^{2\nu} \Gamma(\nu + \frac{d}{2})}{\pi^{\frac{d}{2}} \Gamma(\nu)} (a^2 + |\boldsymbol{\omega}|^2)^{-\nu - \frac{d}{2}}, \quad \boldsymbol{\omega} \in \mathbb{R}^d.$$

*We can easily find out that*

$$\left| \delta^{d-(2\nu+d)} \bar{f}_\delta^\tau(\boldsymbol{\omega}; \sigma^2, \nu, a) - g(\boldsymbol{\omega}; \sigma^2, \nu, a) \right| = O(\max\{\delta^{4\tau - 2\nu - d}, \delta^2\})$$

*where*

$$g(\boldsymbol{\omega}; \sigma^2, \nu, a) = \frac{\sigma^2 a^{2\nu} \Gamma(\nu + \frac{d}{2})}{\pi^{\frac{d}{2}} \Gamma(\nu)} \left\{ \sum_{j=1}^{d} 4 \sin^2 \left( \frac{\omega_j}{2} \right) \right\}^{2\tau} \sum_{\mathbf{Q} \in \mathbb{Z}^d} |\boldsymbol{\omega} + 2\pi \mathbf{Q}|^{-(2\nu+d)}.$$

*We can reparametrize* $(\sigma^2, \nu, a)$ *into* $(c, \alpha, a)$ *with*

$$c = \frac{\sigma^2 a^{2\nu} \Gamma(\nu + \frac{d}{2})}{\pi^{\frac{d}{2}} \Gamma(\nu)},$$

$$\alpha = 2\nu + d.$$

*Then, $g(\boldsymbol{\omega}; \sigma^2, \nu, a)$ can be rewritten as $g(\boldsymbol{\omega}; c, \alpha)$. It can be shown that the assumption (A2) is met with $\gamma = 2$ whenever $4\tau - \alpha \geq 2$. Taking $\mathcal{A}$ as a compact subset of $(d, 4\tau - 2]$, it can also be shown that $g(\boldsymbol{\omega}; c, \alpha)$ satisfies the rest of Assumption 1. Now Theorem 5.1 is applicable so that our estimator minimizing $L_N$ is consistent with*

$$N^{\min\{\frac{d}{2},2\}}(\hat{\alpha} - \alpha_0) = O_P(1),$$
$$N^{\min\{\frac{d}{2},2\}}(\log N)^{-1}(\hat{c} - c_0) = N^{\min\{\frac{d}{2},2\}}(\hat{\alpha} - \alpha_0) + o_P(1),$$

*when the true parameter is $(c_0, \alpha_0)$. Moreover, their asymptotic distributions can also be derived when $d \leq 3$.*

*There are several works which estimate parameters in the isotropic Matérn covariance model. Zhang (2004) proved that under the nested sampling fixed domain asymptotic scheme, $\sigma^2 a^{2\nu}$ can be consistently estimated using the maximum likelihood estimation method assuming $\nu$ is known and $d \leq 3$. Here, the nested sampling scheme means that the next sample is obtained by adding more observations to the previous sample. Speaking in detail, what Zhang (2004) actually does is to maximize the likelihood with respect to $\sigma^2$ after fixing $a$ as an arbitrary value $a_1 > 0$. It is shown that the corresponding estimator, $\hat{\sigma}^2 a_1^{2\nu}$, converges to $\sigma_0^2 a_0^{2\nu}$ almost surely when the true parameter for $(\sigma^2, a)$ is $(\sigma_0^2, a_0)$. Kaufman et al. (2008) showed that under the same assumptions as Zhang (2004), maximizing tapered likelihood gives asymptotically the same result. Du et al. (2009) derived the asymptotic normality for both MLE and tapered MLE when $d = 1$ and $\nu$ is known under the nested sampling scheme. Specifically, writing the sample size as $N$, what Du et al. (2009) showed is*

*that* $N^{\frac{1}{2}}(\hat{\sigma}^2 a_1^{2\nu} - \sigma_0^2 a_0^{2\nu})$ *converges weakly to a normal distribution when* $\hat{\sigma}^2$ *is achieved through both ML and tapered ML method. Wang and Loh (2011) extended the result of Du et al. (2009) to* $d \leq 3$ *but still* $\nu$ *is assumed to be known. Summarizing, we can see that* $\sigma^2 a^{2\nu}$ *can be estimated consistently through both MLE and tapered MLE and the estimator has asymptotic normality under the scaling by square root of the sample size.*

*Compared to the aforementioned results, our estimator has both similar and different aspects with them. Our method jointly estimates tail parameter* $c$ *and* $\alpha$*, which corresponds to joint estimation of* $\sigma^2 a^{2\nu}$ *and* $\nu$*. In other words,* $\nu$ *need not be known and can be estimated simultaneously with* $\sigma^2 a^{2\nu}$*. Moreover,* $N^{\frac{d}{2}}(\hat{\nu} - \nu_0)$ *and* $N^{\frac{d}{2}}(\log N)^{-1}(\widehat{\sigma^2 a^{2\nu}} - \sigma_0^2 a_0^{2\nu_0})$ *converges weakly to a normal distribution when the true parameter is* $(\sigma_0^2, \nu_0, a_0)$*. Since we assume lattice observation, we can understand* $N^{\frac{d}{2}}$ *as a square root of the sample size. At first glance, the asymptotic order for* $\widehat{\sigma^2 a^{2\nu}}$ *seems inferior by the presence of additional* $\log N$ *term. However this is not an appropriate comparison in the point of view that this* $\log N$ *amount of sacrifice arises from assuming* $\nu$ *to be unknown. In fact, it can be shown that the order of convergence for* $\widehat{\sigma^2 a^{2\nu}}$ *is recovered to* $O_P(N^{-\frac{d}{2}})$ *if we assume* $\nu$ *to be known. It is worthwhile to mention that our result holds even when the underlying structure does not actually follow a Matérn structure. This is because we only assume the tail structure of spectral density. For further explanation with an example, see Example 5.4.*

*There is another work, Loh et al. (2021), which estimates both* $\sigma^2 a^{2\nu}$ *and* $\nu$ *in the isotropic Matérn class when* $d \leq 3$*. They used higher-order quadratic variations to estimate* $\nu$ *first and plugged it in to achieve the estimator for*

$\sigma^2 a^{2\nu}$. *Three different types of sampling schemes are considered together in Loh et al. (2021), namely stratified sampling designs, randomized sampling designs and deformed lattice designs. Lattice sampling design can be seen as a special case of stratified sampling design. In the method proposed in Loh et al. (2021), their estimator for* $\nu$ *under stratified sampling design satisfies* $\mathbb{E}|\hat{\nu} - \nu_0| = O(N^{-\frac{d}{d+2}})$ *and the corresponding estimator for* $\sigma^2 a^{2\nu}$ *satisfies* $\mathbb{E}|\log(\widehat{\sigma^2 a^{2\nu}}) - \log(\sigma_0^2 a_0^{2\nu_0})| = O(N^{-\frac{d}{d+2}} \log N)$. *It can be inferred from their proof that these orders will be improved when especially lattice sampling design is assumed, however we do not verify this explicitly. Although Loh et al. (2021) cannot be directly compared with our method, we mention as a reference that the estimator from our method satisfy* $\hat{\nu} - \nu_0 = O_P(N^{-\frac{d}{2}})$ *and* $\log(\widehat{\sigma^2 a^{2\nu}}) - \log(\sigma_0^2 a_0^{2\nu_0}) = O_P(N^{-\frac{d}{2}} \log N)$ *when* $d \leq 3$. *We also point out that our method is also applicable for* $d \geq 4$ *and for non-Matérn cases as long as the tail structure coincides. (See Example 5.4.)*

**Example 5.3.** *Another example is the anisotropic Matérn model whose spectral density is*

$$f(\boldsymbol{\omega}; \sigma^2, \nu, A, a) = \frac{\sigma^2 a^{2\nu} \Gamma(\nu + \frac{d}{2})}{\pi^{\frac{d}{2}} \Gamma(\nu)} (a^2 + |A^T \boldsymbol{\omega}|^2)^{-\nu - \frac{d}{2}}, \quad \boldsymbol{\omega} \in \mathbb{R}^d,$$

*where* $A$ *is an anisotropy matrix. It is easy to verify that the model becomes identifiable when* $A$ *is assumed to be upper triangular matrix whose determinant is 1 (Anderes, 2010). The same argument with the isotropic Matérn also applies to the anisotropic Matérn class, from which (A2) also holds with* $\gamma = 2$. *Applying our method, simultaneously consistent estimator for* $(c, \alpha, A)$ *is achieved with its asymptotic order and the asymptotic distribution is only provided when*

$d \leq 3$.

*Anderes (2010) also proposed estimators of $\sigma^2 a^{2\nu}$ and $A$ for the anisotropic Matérn model. The similarity of Anderes (2010) to our result is that both assume lattice observations and are applicable for general dimension $d$. However, Anderes (2010) neither provided the convergence rate directly nor derived the asymptotic distribution while our result contains both. Another difference is that Anderes (2010) assumed $\nu$ to be known while our method can estimate all the $\sigma^2 a^{2\nu}, \nu$ and $A$, simultaneously. There is one more critical difference, which is that Anderes (2010) proved that the method therein can separately estimate $\sigma^2$ and $a$ when $d \geq 5$. This separate estimation cannot be conducted with our method in the current chapter, since the design of the objective function $L_N$ is not capable of distinguishing $\sigma^2$ and $a$ in the tail behavior point of view. However, using the approach in the next chapter which is modified for parametric models, we can also prove that $\sigma^2$ and $a$ can be separately estimated when $d \geq 5$ and the result coincides with Anderes (2010).*

**Example 5.4.** *The generalized Wendland covariance class is designed for a compactly supported covariance structure, whose fixed domain asymptotic result can be found in Bevilacqua et al. (2019). Generalized Wendland covariance function is parametrized with 4 parameters, $\nu, \kappa, \beta$ and $\sigma^2$. For $\kappa, \sigma^2, \beta > 0$ and $\mu \geq \lambda = \frac{d+1}{2} + \kappa$, its spectral density satisfies*

$$f(\boldsymbol{\omega}) = \sigma^2 L^{\varsigma} \beta^d \Big[ c_3^{\varsigma} (\beta|\boldsymbol{\omega}|)^{-2\lambda} \left\{ 1 + O(|\boldsymbol{\omega}|^{-2}) \right\} + c_4^{\varsigma} (\beta|\boldsymbol{\omega}|)^{-(\mu+\lambda)} \left\{ \cos(\beta|\boldsymbol{\omega}| - c_5^{\varsigma}) + O(|\boldsymbol{\omega}|^{-1}) \right\} \Big]$$

*for $|\boldsymbol{\omega}| \to \infty$ where $c_3^{\varsigma} = \frac{\Gamma(\mu+2\lambda)}{\Gamma(\mu)}, c_4^{\varsigma} = \frac{\Gamma(\mu+2\lambda)}{\Gamma(\lambda) 2^{\lambda-1}}, c_5^{\varsigma} = \frac{\pi}{2}(\mu+\lambda), L^{\varsigma} = \frac{K^{\varsigma}\Gamma(\kappa)}{2^{1-\kappa} B(2\kappa, \mu+1)}$*

*and*

$$K^{\varsigma} = \frac{2^{-\kappa-d+1}\pi^{-\frac{d}{2}}\Gamma(\mu+1)\Gamma(2\kappa+d)}{\Gamma(\kappa+\frac{d}{2})\Gamma(\mu+2\lambda)}$$

*where* $\boldsymbol{\varsigma} = (\mu, \kappa, d)$. $f(\boldsymbol{\omega})$ *can be similarly extended to the case* $\kappa = 0$. *(See Theorem 1 in Bevilacqua et al. (2019).) For convenience, we only state for* $\kappa > 0$ *but the same holds for* $\kappa = 0$. *From the above, we have*

$$\begin{aligned}
\bar{f}_{\delta}^{\tau}(\boldsymbol{\omega}) =& \frac{\sigma^2 L^{\varsigma}}{\delta^d}\left\{\sum_{j=1}^{d} 4\sin^2\left(\frac{\omega_j}{2}\right)\right\}^{2\tau}\Bigg[c_3^{\varsigma}\beta^{-1-2\kappa}\delta^{2\lambda}\sum_{\mathbf{Q}\in\mathbb{Z}^d}|\boldsymbol{\omega}+2\pi\mathbf{Q}|^{-2\lambda} \\
&+O(1)\times c_3^{\varsigma}\beta^{-1-2\kappa}\delta^{2\lambda+2}\sum_{\mathbf{Q}\in\mathbb{Z}^d}|\boldsymbol{\omega}+2\pi\mathbf{Q}|^{-2\lambda-2} \\
&+c_4^{\varsigma}\beta^{-1-2\kappa-(\mu-\lambda)}\delta^{\mu+\lambda}\sum_{\mathbf{Q}\in\mathbb{Z}^d}|\boldsymbol{\omega}+2\pi\mathbf{Q}|^{-(\mu+\lambda)}\cos\left(\frac{\beta|\boldsymbol{\omega}+2\pi\mathbf{Q}|}{\delta}-c_5^{\varsigma}\right) \\
&+O(1)\times c_4^{\varsigma}\beta^{-1-2\kappa-(\mu-\lambda)}\delta^{\mu+\lambda+1}\sum_{\mathbf{Q}\in\mathbb{Z}^d}|\boldsymbol{\omega}+2\pi\mathbf{Q}|^{-(\mu+\lambda+1)}\Bigg].
\end{aligned}$$

*Define* $g(\boldsymbol{\omega}; c, \lambda)$ *as*

$$g(\boldsymbol{\omega}; c, \lambda) = c\left\{\sum_{j=1}^{d} 4\sin^2\left(\frac{\omega_j}{2}\right)\right\}^{2\tau}\sum_{\mathbf{Q}\in\mathbb{Z}^d}|\boldsymbol{\omega}+2\pi\mathbf{Q}|^{-2\lambda}$$

*with* $g(\mathbf{0}; c, \lambda) = 0$. *If* $\mu > \lambda$ *and* $4\tau > 2\lambda + 2$, *then we have*

$$\sup_{\boldsymbol{\omega}\in\mathbb{T}^d}\left|\delta^{d-2\lambda}\bar{f}_{\delta}^{\tau}(\boldsymbol{\omega}) - g(\boldsymbol{\omega}; c, \lambda)\right| = O(\delta^{\min\{2,\mu-\lambda\}})$$

*where* $c = \sigma^2\beta^{-1-2\kappa}L^{\varsigma}c_3^{\varsigma}$. *Moreover, under the same restriction on parameters, it can be shown using the known fact about the derivative of generalized hypergeometric function that all the first partial derivatives of* $\delta^{d-\alpha}\bar{f}_{\delta}^{\tau}(\boldsymbol{\omega})$ *converge uniformly. The rest of Assumption 1 holds because the limit function* $g$ *is identical to the Matérn case, therefore we can apply Theorem 5.1. Not only the*

*estimator* $(\hat{c}, \hat{\lambda})$ *is consistent but the limiting distribution can be derived when* $\frac{d}{2} < \min\{2, \mu - \lambda\}$. *Note that this is especially simplified as* $\mu > \lambda + \frac{d}{2}$ *when* $d \leq 3$.

*In Bevilacqua et al. (2019), they assumed known* $\mu$ *and* $\kappa$ *(then* $\lambda$ *is also known and thus* $c_3^{\varsigma}$ *and* $L^{\varsigma}$ *are all known),* $\mu > \lambda + \frac{d}{2}$ *and* $d \leq 3$. *Then, they proved that maximizing the likelihood function with respect to* $\sigma^2$ *while* $\beta$ *is fixed as an arbitrary value gives a consistent estimator for* $\sigma^2\beta^{-1-2\kappa}$ *which is a similar result as in Zhang (2004) for the Matérn class. Using our method, we can estimate* $\lambda$ *and* $\sigma^2\beta^{-1-2\kappa}L^{\varsigma}c_3^{\varsigma}$ *both consistently. Comparisons for the asymptotic order and normality when* $d \leq 3$ *are similar to that for the Matérn class. Similarly again, our method has a benefit in that it can be applied for data from general dimension* $d$, *but has a limit in that our method is only for lattice data.*

*In this example, we discuss about the benefit of the tail behavior assumption when the covariance model is misspecified. Observe that by reparametrizing* $2\lambda$ *as* $\alpha$, *the limiting function* $g(\boldsymbol{\omega}; c, \alpha)$ *becomes exactly the same with the isotropic Matérn case. This has already been expected from Theorem 5 in Bevilacqua et al. (2019), in which Gaussian measure equivalences between generalized Wendland class and Matérn class are proved by showing their spectral densities share same tail behaviors. Zhang (2004) and Bevilacqua et al. (2019) both used likelihoods which depend on which parametric model has been assumed. Hence, a question such as what if we use Matérn class likelihood function when the true underlying structure follows in fact generalized Wendland class and vice versa has not been answered. Our method serves an answer for this question. Using* $L_N$ *con-*

*structed with the same limit function $g(\boldsymbol{\omega}; c, \alpha)$, both Matérn and generalized Wendland class parameters can be estimated. The estimated parameter $(\hat{c}, \hat{\alpha})$ would be an estimator for $\left(\frac{\sigma^2 a^{2\nu}\Gamma(\nu+\frac{d}{2})}{\pi^{\frac{d}{2}}\Gamma(\nu)}, 2\nu + d\right)$ when the underlying covariance structure were Matérn, and for $\left(\sigma^2\beta^{-1-2\kappa}L^{\varsigma}c_3^{\varsigma}, 2\lambda\right)$ when the underlying covariance structure were generalized Wendland. It does not matter what the real underlying model was because two models induce equivalent Gaussian measures on any bounded domain, as proved in Theorem 5 in Bevilacqua et al. (2019).*

# Chapter 6

# Tail Parameter Estimation in parametric covariance models

The tail parameter estimation method proposed in Chapter 5 requires evaluation of $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ and $g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})$ at $\boldsymbol{\omega} \in \Omega_M$. When $\hat{\mathcal{I}}_{N,\delta,M}$ is the tapered periodogram or the smoothed periodogram with a compactly supported kernel whose Fourier coefficients are completely known and a bandwidth $h = \frac{\pi}{M}$, it can be exactly evaluated as explained in Chapter 4. For evaluating $g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})$, however, we cannot help but approximating this value, since $g(\boldsymbol{\omega};\boldsymbol{\theta})$ is defined with the infinite sum whose value cannot be exactly obtained. Fortunately, we can modify our objective function so that it can be evaluated exactly when we assume a parametric covariance model instead of only a tail behavior model on the spectral density.

Assume that we have a parametric covariance function model $C(\mathbf{s};\boldsymbol{\theta},\boldsymbol{\eta}), \mathbf{s} \in \mathbb{R}^d$ with the corresponding spectral density $f(\boldsymbol{\omega};\boldsymbol{\theta},\boldsymbol{\eta}), \boldsymbol{\omega} \in \mathbb{R}^d$ so that

$\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega};\boldsymbol{\theta},\boldsymbol{\eta}) \to g(\boldsymbol{\omega};\boldsymbol{\theta})$ as $\delta \downarrow 0$. We can modify $L_N(\boldsymbol{\theta})$ into $R_N(\boldsymbol{\theta},\boldsymbol{\eta})$ defined as

$$R_N(\boldsymbol{\theta},\boldsymbol{\eta}) = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \left\{ \frac{\hat{\mathcal{I}}_{N,\delta,M}(\boldsymbol{\omega})}{\bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega};\boldsymbol{\theta},\boldsymbol{\eta})} + \log \bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega};\boldsymbol{\theta},\boldsymbol{\eta}) \right\} \tag{6.1}$$

where $\bar{f}^{\tau}_{N,\delta,M}(\cdot;\boldsymbol{\theta},\boldsymbol{\eta}) = \bar{f}^{\tau}_{\delta}(\cdot;\boldsymbol{\theta},\boldsymbol{\eta}) * K_N * K_M$ for the tapered periodogram case and $\bar{f}^{\tau}_{N,\delta,M}(\cdot;\boldsymbol{\theta},\boldsymbol{\eta}) = \bar{f}^{\tau}_{\delta}(\cdot;\boldsymbol{\theta},\boldsymbol{\eta}) * K_N * \frac{1}{h^d}K\left(\frac{\cdot}{h}\right)$ for the smoothed periodogram with a compactly supported kernel. Observe that $\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega};\boldsymbol{\theta},\boldsymbol{\eta})$ is the spectral density of $Y^{\tau}_{\delta}(\mathbf{J}), \mathbf{J}\in\mathbb{Z}^d$ when the underlying covariance function of the original process $Z(\mathbf{s}), \mathbf{s}\in\mathbb{R}^d$ is $C(\mathbf{s};\boldsymbol{\theta},\boldsymbol{\eta})$. From the definition of $Y^{\tau}_{\delta}(\mathbf{J})$, we can derive its covariance function $C^{\tau}_{\delta}(\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta})$ which is expressed by a finite linear combination of $C(\delta\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta})$. Specifically, it can be shown that

$$C^{\tau}_{\delta}(\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta}) = \Delta^{2\tau} C_{\delta}(\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta}),$$
$$C_{\delta}(\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta}) = C(\delta\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta})$$

where $\Delta$ is the discrete Laplacian operator defined in Chapter 3. Then we have a relation

$$\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega};\boldsymbol{\theta},\boldsymbol{\eta}) = \frac{1}{(2\pi)^d} \sum_{\mathbf{J}\in\mathbb{Z}^d} C^{\tau}_{\delta}(\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta}) e^{-i\langle\boldsymbol{\omega},\mathbf{J}\rangle}.$$

Convolution of this with $K_N$ is given as

$$\begin{aligned}\bar{f}^{\tau}_{N,\delta}(\boldsymbol{\omega};\boldsymbol{\theta},\boldsymbol{\eta}) &= \frac{1}{(2\pi)^{2d}} \sum_{\mathbf{J}\in\mathbb{Z}^d} \sum_{\mathbf{K}:\|\mathbf{K}\|_\infty\le N} a_{N-|\mathbf{K}|} C^{\tau}_{\delta}(\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta}) \int_{\mathbb{T}^d} e^{-i\langle\boldsymbol{\upsilon},\mathbf{J}\rangle} e^{-i\langle\boldsymbol{\omega}-\boldsymbol{\upsilon},\mathbf{K}\rangle} \mathbf{d}\boldsymbol{\upsilon} \\ &= \frac{1}{(2\pi)^d} \sum_{\|\mathbf{J}\|_\infty\le N} a_{N-|\mathbf{J}|} C^{\tau}_{\delta}(\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta}) e^{-i\langle\boldsymbol{\omega},\mathbf{J}\rangle}\end{aligned}$$

where $a_{N-|\mathbf{J}|}$ is defined as in Chapter 4. Now, if we use the tapered periodogram for $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}$, then $\bar{f}^{\tau}_{N,\delta,M} = \bar{f}^{\tau}_{N,\delta} * K_M$ where $K_M$ is the Fejér kernel of an order

$M$ so that it becomes

$$\bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega};\boldsymbol{\theta},\boldsymbol{\eta}) = \frac{1}{(2\pi)^d} \sum_{\|\mathbf{J}\|_\infty \le M} a_{N-|\mathbf{J}|} a_{M-|\mathbf{J}|} C^{\tau}_{\delta}(\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta}) e^{-i\langle \boldsymbol{\omega},\mathbf{J}\rangle}.$$

When $\hat{\mathcal{I}}^{\tau}_{N,\delta,M}$ is the smoothed periodogram with a kernel $K$ and a bandwidth $h = \frac{\pi}{M}$, then $\bar{f}^{\tau}_{N,\delta,M} = \bar{f}^{\tau}_{N,\delta} * \frac{1}{h^d} K\left(\frac{\cdot}{h}\right)$ so that it becomes

$$\bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega};\boldsymbol{\theta},\boldsymbol{\eta}) = \frac{1}{(2\pi)^d} \sum_{\|\mathbf{J}\|_\infty \le N} a_{N-|\mathbf{J}|} \hat{K}(h\mathbf{J}) \hat{C}^{\tau}_{\delta}(\mathbf{J};\boldsymbol{\theta},\boldsymbol{\eta}) e^{-i\langle \boldsymbol{\omega},\mathbf{J}\rangle}$$

where

$$\hat{K}(\mathbf{s}) = \int_{\mathbb{T}^d} K(\mathbf{x}) e^{-i\langle \mathbf{x},\mathbf{s}\rangle} \mathbf{dx}.$$

Since these are finite sums, $R_N(\boldsymbol{\theta},\boldsymbol{\eta})$ can be evaluated without approximation. We propose two ways to estimate $\boldsymbol{\theta}$ using $R_N(\boldsymbol{\theta},\boldsymbol{\eta})$. The first approach is to fix $\boldsymbol{\eta}$ as an arbitrary value $\boldsymbol{\eta}_1$ and minimize $R_N(\boldsymbol{\theta},\boldsymbol{\eta}_1)$ with respect to $\boldsymbol{\theta}$. The second approach is to minimize $R_N(\boldsymbol{\theta},\boldsymbol{\eta})$ with respect to both $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. The following additional assumption is helpful to justify these attempts theoretically.

**Assumption 2.**

*(A7) The parameter space $\mathcal{H}$ for $\boldsymbol{\eta}$ is compact. Moreover, the following uniform convergences hold:*

$$\sup_{\boldsymbol{\omega}\in\mathbb{T}^d,\alpha\in\mathcal{A},\boldsymbol{\beta}\in\mathcal{B},\boldsymbol{\eta}\in\mathcal{H}} \left|\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta},\boldsymbol{\eta}) - g(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})\right| \to 0,$$

$$\sup_{\boldsymbol{\omega}\in\mathbb{T}^d,\alpha\in\mathcal{A},\boldsymbol{\beta}\in\mathcal{B},\boldsymbol{\eta}\in\mathcal{H}} \left\|\frac{\partial}{\partial(\alpha,\boldsymbol{\beta})}\left(\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta},\boldsymbol{\eta}) - g(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})\right)\right\|_\infty \to 0,$$

$$\sup_{\boldsymbol{\omega}\in\mathbb{T}^d,\alpha\in\mathcal{A},\boldsymbol{\beta}\in\mathcal{B},\boldsymbol{\eta}\in\mathcal{H}} \left\|\frac{\partial^2}{\partial(\alpha,\boldsymbol{\beta})\partial(\alpha,\boldsymbol{\beta})^T}\left(\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta},\boldsymbol{\eta}) - g(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})\right)\right\|_\infty \to 0.$$

With this additional assumption, the following theorem says that in either way we can get a consistent estimator of $\boldsymbol{\theta}$ with desirable fixed domain asymptotic properties.

**Theorem 6.1.** *Assume the same as in Theorem 5.1 with Assumption 1 and additionally Assumption 2. Suppose $R_N(\boldsymbol{\theta}, \boldsymbol{\eta})$ is defined as in* (6.1).

*If we fix $\boldsymbol{\eta}$ as arbitrary value, say $\boldsymbol{\eta}_1 \in \mathcal{H}$, then the minimizer $\hat{\boldsymbol{\theta}}$ of $R_N(\boldsymbol{\theta}, \boldsymbol{\eta}_1)$ converges to $\boldsymbol{\theta}$ in probability and its asymptotic properties are the same as that achieved in Theorem 5.1.*

*If we fully minimize $R_N(\boldsymbol{\theta}, \boldsymbol{\eta})$ with respect to $(\boldsymbol{\theta}, \boldsymbol{\eta}) \in \Theta \times \mathcal{H}$, then $\hat{\boldsymbol{\theta}}$ of the minimizer $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$ converges to $\boldsymbol{\theta}$ in probability and its asymptotic properties are same as that achieved in Theorem 5.1.*

*Proof.* The proof is similar to Theorem 5.1, hence we briefly sketch the proof omitting details. First we define

$$R_{0N}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega} \in \Omega_M} \left\{ \frac{\delta^{d-\alpha_0} \hat{\mathcal{I}}_{N,\delta,M}(\boldsymbol{\omega})}{\delta^{d-\alpha} \bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\eta})} + \log\left(\delta^{d-\alpha} \bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\eta})\right) \right\}.$$

Note that $R_{0N}(\boldsymbol{\theta}, \boldsymbol{\eta})$ is close to $L_0(\boldsymbol{\theta})$ where $L_0$ is defined in the proof of Proposition B.3. Using almost same arguments with Proposition B.3, consistency of $\hat{\boldsymbol{\theta}}$ for both cases can be shown.

To show the limiting distribution of $\hat{\boldsymbol{\theta}}$, first we consider the case when $\boldsymbol{\eta}$ is fixed as $\boldsymbol{\eta}_1$. In this case, almost same arguments with Proposition B.4 shows the desired result. It is more complicated when $\boldsymbol{\eta}$ is not fixed. In this case, all we need to show is that $\frac{\partial}{\partial \boldsymbol{\theta}} R_{0N}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) = \mathbf{0}$. Then as in the former case, we can just follow the arguments in Proposition B.4 again. Since $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$ is the minimizer of

$R_{0N}(\boldsymbol{\theta}, \boldsymbol{\eta})$ on $\Theta \times \mathcal{H}$, we have

$$\left\langle \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\theta}} R_{0N}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) \\ \frac{\partial}{\partial \boldsymbol{\eta}} R_{0N}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix} \right\rangle \geq 0$$

for all feasible $(\mathbf{v}^T, \mathbf{w}^T)^T$. When we can say that $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$ are in the interior of $\Theta$ and $\mathcal{H}$ respectively, any vector in the Euclidean space of the corresponding dimension is feasible for $\mathbf{v}$ and $\mathbf{w}$. Hence we get both $\frac{\partial}{\partial \boldsymbol{\theta}} R_{0N}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) = \mathbf{0}$ and $\frac{\partial}{\partial \boldsymbol{\eta}} R_{0N}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) = \mathbf{0}$. When $\hat{\boldsymbol{\eta}}$ is on the boundary of $\mathcal{H}$, not all vectors are feasible for $\mathbf{w}$. However we know that $\hat{\boldsymbol{\theta}}$ converges to its true value $\boldsymbol{\theta}_0$ which lies in the interior of $\Theta$, hence so does $\hat{\boldsymbol{\theta}}$ with probability tending to 1. Thus any vector is feasible for $\mathbf{v}$. This means that although we cannot say whether $\frac{\partial}{\partial \boldsymbol{\eta}} R_{0N}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$ is a zero vector or not, we still have $\frac{\partial}{\partial \boldsymbol{\theta}} R_{0N}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) = \mathbf{0}$. With this we can show the asymptotic order and normality of $\hat{\boldsymbol{\theta}}$ when $\boldsymbol{\eta}$ is not fixed. Note that we require the assumption (A7) to deal with the limiting distribution of $\frac{\partial}{\partial \boldsymbol{\theta}} R_{0N}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_1)$ when $\boldsymbol{\eta}$ is fixed as $\boldsymbol{\eta}_1$ and that of $\frac{\partial}{\partial \boldsymbol{\theta}} R_{0N}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}})$ when $\boldsymbol{\eta}$ is not fixed.

Now the argument from $R_{0N}$ to $R_N$ is the same as that from $L_{0N}$ to $L_N$ in Theorem 5.1. □

**Remark 6.2.** *The approach in Theorem 6.1 is actually not limited to the estimation within a certain parametric class. It can be also used to enhance the estimation accuracy when the underlying covariance structure is unknown. Here we explain how this works through an example. Suppose that the spectral density $f$ satisfies $f(\boldsymbol{\omega}) \sim c|\boldsymbol{\omega}|^{-\alpha}$ as $|\boldsymbol{\omega}| \to \infty$. This tail behavior model is the same as an isotropic Matérn model. In the real application, we never know whether (A2) in Assumption 1 and (A6) in Assumption 2 holds or not. Hence we need to*

*assume that both hold with some* $\gamma > 0$. *Now consider to construct* $R_N(c, \alpha, a)$ *in which* $\bar{f}^{\tau}_{N,\delta,M}(c, \alpha, a)$ *is from the Matérn model. Modifying the proof slightly, it can be shown that by minimizing this* $R_N(c, \alpha, a)$ *over* $c \in \mathbb{R}^+$, $\alpha$ *in some compact interval containing the true value, and* $a$ *being fixed or in any compact interval, the minimizer* $\hat{c}$ *and* $\hat{\alpha}$ *become consistent estimators for the tail parameters of the underlying spectral density whose asymptotic orders become*

$$N^{\min\{\frac{d}{2},2,\gamma\}}(\hat{\alpha} - \alpha_0) = O_P(1),$$
$$N^{\min\{\frac{d}{2},2,\gamma\}}(\log N)^{-1}(\hat{c} - c_0) = O_P(1),$$

*where* $c_0$ *and* $\alpha_0$ *indicate the true values. Moreover, when* $\frac{d}{2} < \min\{2, \gamma\}$, *the asymptotic distribution of the estimator can be derived as same as in Theorems 5.1 and 6.1.*

*For a more specific example, this can be used when the underlying covariance structure is the generalized Wendland model. In Example 5.4, it has been already discussed that* $\gamma$ *in (A2) is* $\min\{2, \mu - \lambda\}$ *under the generalized Wendland model. When* $\mu - \lambda > \frac{d}{2}$ *and* $d \leq 3$, *then* $\min\{\frac{d}{2}, 2, \gamma\} = \frac{d}{2}$. *This means that when* $d \leq 3$, $R_N(c, \alpha, a)$ *constructed by the Matérn covariance function gives a consistent and asymptotically normal estimator when the underlying covariance structure is the generalized Wendland model. Of course, this argument would also hold converting the role of two covariance models, i.e., the underlying structure is the Matérn model and* $R_N$ *is constructed by the generalized Wendland model.*

# Chapter 7

# Simulation Studies

To illustrate our theoretical result, we conducted experiments using generated data. The first experiment uses Gaussian stationary random fields $Z$ on $\mathbb{R}^2$ with isotropic Matérn covariance structures to generate samples for $\{Y_\delta^\tau(\mathbf{J}) : \mathbf{J} \in \mathbb{Z}_N^2\}$. Aware that this requires a sample $Z(\mathbf{s})$ at $\mathbf{s} \in \left\{-\frac{\tau}{N}, -\frac{\tau-1}{N}, \cdots, \frac{N-1+\tau}{N}\right\}^2$ which is composed of observations at $(N+2\tau)^2$ locations. This $\tau$-sized buffer on each boundary emerges because of our setting $N\delta = 1$ imposed for the sake of convenience. We briefly mention that this buffer is unnecessary both in theory and application if $N\delta$ is fixed as $b$, allowing $b$ to have different value than 1. Recall that the isotropic Matérn covariance function is given as

$$C(\mathbf{s}; \sigma^2, \nu, a) = \frac{\sigma^2 (a|\mathbf{s}|)^\nu}{\Gamma(\nu) 2^{\nu-1}} \mathcal{K}_\nu(a|\mathbf{s}|)$$

with spectral density

$$f(\boldsymbol{\omega}; \sigma^2, \nu, a) \sim \frac{\sigma^2 a^{2\nu} \Gamma(\nu + d/2)}{\pi^{d/2} \Gamma(\nu)} |\boldsymbol{\omega}|^{-2\nu-d} = c|\boldsymbol{\omega}|^{-\alpha}.$$

In this model we are interested in estimating tail parameter $\boldsymbol{\theta} = (c, \alpha)$. For simulation, we fixed $\sigma^2 = 1$ and considered four different models: $(\nu, a) = (0.5, 2.1), (0.5, 9), (1.5, 5)$ and $(1.5, 14.3)$. Smoothness parameter $\nu = 0.5$ and 1.5 are chosen as representatives for a continuous but nondifferentiable random field and once but not twice differentiable random field in a mean-square sense, respectively. For $\nu = 0.5$, $a = 2.1$ and $a = 9$ are taken to illustrate relatively strong and weak dependence. These values for the scale parameter are calculated by seeking $a$ that makes $C(\mathbf{s}; 1, 0.5, a)$ close to 0.5 and 0.05 respectively when $|\mathbf{s}| = \frac{1}{3}$. $a = 5$ and $a = 14.3$ for $\nu = 1.5$ are taken similarly. For each model, we considered $N = 50$ and $N = 100$ to demonstrate convergence properties of our estimators. Since our asymptotic theory requires $4\tau - \alpha \geq 2$ (see Example 5.2) and $d = 2$ so that $\nu = 0.5, 1.5$ implies $\alpha = 2\nu + d = 3, 5$ with $\tau = 2$. For each set of $(\nu, a, N)$, 500 repeated experiments that estimate tail parameters by minimizing $L_N$ given in Theorem 5.1 are made and the results are drawn as histograms.

Before delving into the result, we would like to mention about the term $g_{N,M}(\boldsymbol{\omega}; \boldsymbol{\theta})$ in $L_N$ which needs to be approximately calculated. Consider first $L_N$ with the tapered periodogram. Observe that $g_{N,M}(\boldsymbol{\theta})$ can be written as

$$
\begin{aligned}
g_{N,M}(\boldsymbol{\omega}; \boldsymbol{\theta}) &= \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} g(\boldsymbol{\upsilon}; \boldsymbol{\theta}) K_N(\mathbf{x} - \boldsymbol{\upsilon}) K_M(\boldsymbol{\omega} - \mathbf{x}) \mathbf{d}\boldsymbol{\upsilon}\mathbf{d}\mathbf{x} \\
&= \frac{1}{(2\pi)^{2d}} \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} g(\boldsymbol{\upsilon}; \boldsymbol{\theta}) \sum_{\|\mathbf{J}\|_\infty \leq N} a_{N-|\mathbf{J}|} e^{i\langle \mathbf{x} - \boldsymbol{\upsilon}, \mathbf{J} \rangle} \sum_{\|\mathbf{K}\|_\infty \leq M} a_{M-|\mathbf{K}|} e^{i\langle \boldsymbol{\omega} - \mathbf{x}, \mathbf{K} \rangle} \mathbf{d}\boldsymbol{\upsilon}\mathbf{d}\mathbf{x}.
\end{aligned}
$$

By writting

$$
\hat{g}(\mathbf{J}; \boldsymbol{\theta}) = \int_{\mathbb{T}^d} g(\boldsymbol{\upsilon}; \boldsymbol{\theta}) e^{-i\langle \boldsymbol{\upsilon}, \mathbf{J} \rangle} \mathbf{d}\boldsymbol{\upsilon}, \quad \mathbf{J} \in \mathbb{Z}^d,
$$

we can write

$$g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{2d}} \int_{\mathbb{T}^d} \sum_{\|\mathbf{J}\|_\infty \le N} a_{N-|\mathbf{J}|} \hat{g}(\mathbf{J};\boldsymbol{\theta}) e^{i\langle \mathbf{x},\mathbf{J}\rangle} \sum_{\|\mathbf{K}\|_\infty \le M} a_{M-|\mathbf{K}|} e^{i\langle \boldsymbol{\omega}-\mathbf{x},\mathbf{K}\rangle} \mathbf{dx}$$
$$= \frac{1}{(2\pi)^d} \sum_{\|\mathbf{J}\|_\infty \le M} a_{N-|\mathbf{J}|} a_{M-|\mathbf{J}|} \hat{g}(\mathbf{J};\boldsymbol{\theta}) e^{i\langle \boldsymbol{\omega},\mathbf{J}\rangle}.$$

Hence we can approximate $g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})$ for each $\boldsymbol{\omega} \in \Omega_M$ and $\boldsymbol{\theta} \in \Theta$ if we can approximate Fourier coefficients $\hat{g}(\mathbf{J};\boldsymbol{\theta})$ for $\mathbf{J} \in \mathbb{Z}^d$ satisfying $\|\mathbf{J}\|_\infty \le M$. We can think that Fourier coefficients can be approximated via Riemann sum as

$$\hat{g}(\mathbf{J};\boldsymbol{\theta}) \approx \left(\frac{2\pi}{N}\right)^d \sum_{\mathbf{K} \in \mathbb{Z}_N^d} g\left(\frac{2\pi\mathbf{K}}{N};\boldsymbol{\theta}\right) e^{-i\left\langle \frac{2\pi\mathbf{K}}{N},\mathbf{J}\right\rangle}.$$

This approximation would work well if $M \ll N$, which is satisfied in our case since we fix $M$ while $N$ goes to infinity. Note that the fast Fourier transform algorithm can be used for above approximation. For $L_N$ with smoothed periodograms with a compactly supported kernel $K$ and a bandwidth $h = \frac{\pi}{M}$, the form of $g_{N,M}$ becomes

$$g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}) = \frac{1}{(2\pi)^d} \sum_{\|\mathbf{J}\|_\infty \le N} a_{N-|\mathbf{J}|} \hat{K}(h\mathbf{J}) \hat{g}(\mathbf{J};\boldsymbol{\theta}) e^{i\langle \boldsymbol{\omega},\mathbf{J}\rangle}$$

which can again be approximately evaluated as similar as the tapered periodogram case.

Meanwhile, we also need to approximate $g(\boldsymbol{\omega};\boldsymbol{\theta})$ because it is defined as an infinite lattice sum which does not have an explicit expression in general. Under the spectral tail model (1.1), recall that $g$ is expressed as

$$g(\boldsymbol{\omega};\boldsymbol{\theta}) = c \left\{ \sum_{j=1}^d 4\sin^2\left(\frac{\omega_j}{2}\right) \right\}^{2\tau} \sum_{\mathbf{Q} \in \mathbb{Z}^d} |\boldsymbol{\omega} + 2\pi\mathbf{Q}|^{-\alpha} h\left(\frac{\boldsymbol{\omega} + 2\pi\mathbf{Q}}{|\boldsymbol{\omega} + 2\pi\mathbf{Q}|}; \alpha, \boldsymbol{\beta}\right).$$

Since $h$ is bounded, the absolute value of each term in the summand is comparable to $|\boldsymbol{\omega} + 2\pi\mathbf{Q}|^{-\alpha}$, which decays fast toward zero as $|\mathbf{Q}| \to \infty$. Hence we can approximate $g$ by taking a threshold, say $q$, and summing the terms up to $\|\mathbf{Q}\|_\infty \leq q$. Note that this threshold should be taken carefully depending on $\alpha$. If $\alpha$ is close to $d$, then small $q$ does not approximate $g$ well. In the current isotropic Matérn experiments with $d = 2$, $q = 10$ is taken for $\nu = 1.5$ which corresponds to $\alpha = 5$ and $q = 20$ for $\nu = 0.5$ which corresponds to $\alpha = 3$. Although we do not depict here, experiments with $q = 10$ under $\nu = 0.5$ showed significant biases which does not vanish as $N$ increases.

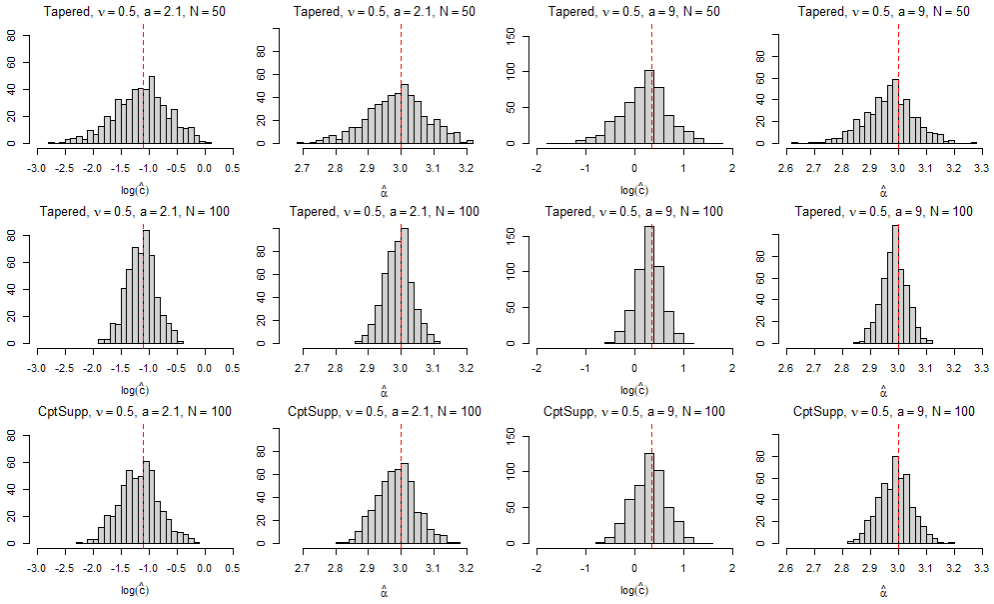


Figure 7.1 Tail parameter estimates of the isotropic Matérn model when $\sigma^2 = 1$ and $\nu = 0.5$. Vertical red line indicates the location of the true parameter.

Figure 7.1 shows how the tail parameters $c$ and $\alpha$ are estimated under four different settings when $\nu = 0.5$. $M$ is set as 10 and $t$ is considered to be

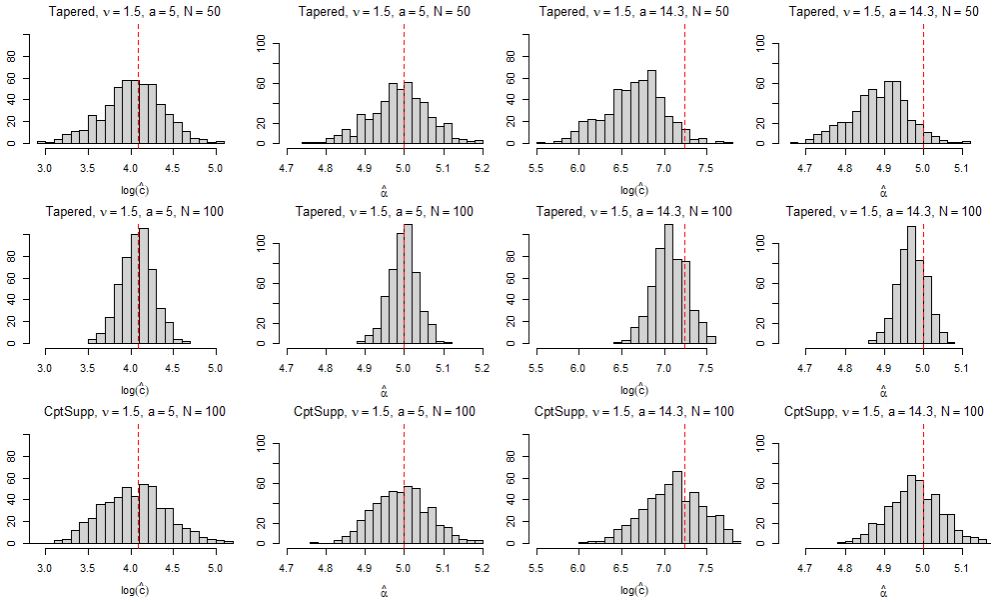


Figure 7.2 Tail parameter estimates of the isotropic Matérn model when $\sigma^2 = 1$ and $\nu = 1.5$. Vertical red line indicates the location of the true parameter.

smaller than $\frac{2\pi}{M}$ for the tapered periodogram and $\frac{2\pi}{M} < t < \frac{4\pi}{M}$ for the smoothed periodogram with a compactly supported kernel. This leads $\Omega_M$ to contain $M^d - 1$ frequencies for the tapered periodogram and $\left(\left[\frac{M}{2}\right] - 1\right) \times M^{d-1} - 3$ frequencies for the smoothed peridogram with a compactly supported kernel. Although it seems that the former makes uses around twice of frequencies than the latter at first glance, there are duplicated information in periodogram originated from its periodicity which makes the amount of frequencies in two cases comparable. The left two columns illustrate estimation result for $\log c$ and $\alpha$ when the dependency is strong and the right two columns correspond to weak dependency. The upper two rows uses the tapered periodogram to construct $L_N$ and the last row uses the smoothed periodogram with a compactly supported kernel. For the compact

| $\nu$ | $a$ | $N$ | Type | $\log c$ | | | $\alpha$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | bias | st. dev. | RMSE | bias | st. dev. | RMSE |
| 0.5 | 2.1 | 50 | tapered | -0.0516 | 0.4887 | 0.4909 | -0.0131 | 0.0924 | 0.0932 |
| | | 100 | tapered | -0.0402 | 0.2539 | 0.2568 | -0.0102 | 0.0425 | 0.0436 |
| | | 100 | cptsupp | -0.0765 | 0.3594 | 0.3671 | -0.0143 | 0.0600 | 0.0616 |
| | 9 | 50 | tapered | -0.1573 | 0.4797 | 0.5044 | -0.0322 | 0.0902 | 0.0957 |
| | | 100 | tapered | -0.0641 | 0.2631 | 0.2705 | -0.0139 | 0.0440 | 0.0461 |
| | | 100 | cptsupp | -0.0593 | 0.3493 | 0.3539 | -0.0111 | 0.0585 | 0.0595 |
| 1.5 | 5 | 50 | tapered | -0.0653 | 0.3562 | 0.3618 | -0.0121 | 0.0741 | 0.0750 |
| | | 100 | tapered | -0.0140 | 0.1917 | 0.1920 | -0.0022 | 0.0347 | 0.0347 |
| | | 100 | cptsupp | -0.0400 | 0.3876 | 0.3893 | -0.0070 | 0.0695 | 0.0698 |
| | 14.3 | 50 | tapered | -0.5907 | 0.3492 | 0.6860 | -0.1121 | 0.0719 | 0.1331 |
| | | 100 | tapered | -0.1783 | 0.1970 | 0.2655 | -0.0299 | 0.0357 | 0.0465 |
| | | 100 | cptsupp | -0.1175 | 0.3726 | 0.3904 | -0.0192 | 0.0671 | 0.0697 |

Table 7.1 Biases, standard deviations and root mean squared errors of estimated $\log c$ and $\alpha$.

support kernel, a multidimensional biweight kernel is used which is defined as

$$K(\mathbf{s}) = \prod_{j=1}^{d} \frac{15}{16}(1 - s_j^2)^2 \mathbb{1}_{[-1,1]}(s_j)$$

where $\mathbb{1}_A(x)$ denotes the indicator function taking values 1 when $x \in A$ and 0 otherwise. The result in the first row corresponds to $N = 50$ while the second and the third row correspond to $N = 100$. For all cases, we can see that both $c$ and $\alpha$ are consistently estimated. Comparing the first two rows clearly shows that estimation variance decreases as $N$ increases which experimentally indicates the consistent tendencies of our estimates. We can also see that the result

of the second row shows the result of lower variance than the third row, implying that using the tapered periodogram gives more accurate estimate than using the smoothed periodogram with the multivariate biweight kernel when $M$ is set as same. Although the tapered periodogram version is beneficial both in estimation accuracy and computational cost, the privilege of using the smoothed periodogram with a compactly supported kernel is that the asymptotic variance of the estimators can be numerically approximated so that statistical inference becomes possible.

Similar experiment for $\nu = 1.5$ is depicted in Figure 7.2. Although all the result shows consistent and asymptotically unbiased tendency as in the former case, the meaningful bias is captured when the spatial dependency is weak ($a = 14.3$) and $N = 50$. This can be explained by the discrepancy between the limiting spectral density and finite sample spectral density. The spectral density of the isotropic Matérn model is of form $f(\boldsymbol{\omega}; c, \alpha, a) = c(a^2 + |\boldsymbol{\omega}|^2)^{-\alpha/2}$ so that

$$\bar{f}_\delta^\tau(\boldsymbol{\omega}; c, \alpha, a) = c\delta^{\alpha-d} \left\{ \sum_{j=1}^{d} 4\sin^2\left(\frac{\omega_j}{2}\right) \right\}^{2\tau} \sum_{\mathbf{Q}\in\mathbb{Z}^d} (a^2\delta^2 + |\boldsymbol{\omega} + 2\pi\mathbf{Q}|^2)^{-\alpha/2}.$$

The term $a^2\delta^2$ becomes ignorable when $N$ is large, however, its effect should be considered under the finite sample situation. Indeed, for our case of $a = 14.3$ with $N = 50$ and $a\delta = 0.286$, the simulation shows that a significant bias was created in the actual estimates. The bias seems to be alleviated when $N$ is increased to 100, which makes sense since $a\delta$ becomes a half of the previous value. We do not depict the result here, but we have observed that this bias almost disappears when $N$ is increased to 200. The results of Figure 7.1 and

7.2 are also summarized in Table 7.1.

Theorem 5.1 implies that when $\frac{d}{2} < \gamma$, which means $d \leq 3$ since $\gamma = 2$ in the Matérn case, the asymptotic behavior of $N^{\frac{d}{2}}(\log N)^{-1}(\log \hat{c} - \log c_0)$ is equal to that of $N^{\frac{d}{2}}(\hat{\alpha} - \alpha_0)$. This means that $\log \hat{c} - \log c_0$ and $\hat{\alpha} - \alpha_0$ are asymptotically perfectly correlated. Figure 7.3 provides an experimental evidence for this theoretical result. We mention here that similar phenomenon has been reported in the simulation study of Stein (1995).

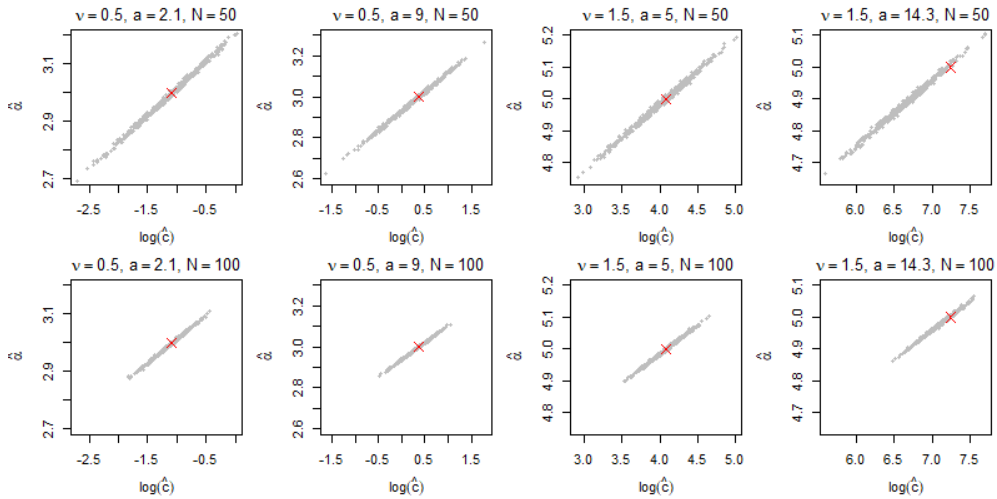


Figure 7.3 Asymptotic perfect linear correlation between $\log \hat{c}$ and $\hat{\alpha}$. Red crosses are locations of true parameters.

We have already discussed that variance of the limiting distribution for the estimated parameters can be evaluated numerically when the smoothed periodogram with a compactly supported kernel is used. This enables us to evaluate approximated confidence intervals for the parameters. For each 500 independent realization of each setting when $N = 100$, we evaluated 95% confidence

intervals. The results are displayed in Figure 7.4. The coverage probabilities calculated from this procedure, which is in Table 7.2, do not become exactly 0.95 as expected. One reason for this is that there is an approximation error in the numerically approximated variance of the limiting distribution. The limiting variance stated in Theorem 5.1 requires evaluations of the terms $g_h$, its derivatives with respect to the tail parameter and $g^2 * \frac{1}{h^{2d}} K^2\left(\frac{\cdot}{h}\right)$. Since $g_h$ cannot be exactly evaluated, we use $g_{L,h} = g * K_L * \frac{1}{h^d} K\left(\frac{\cdot}{h}\right)$ for some large $L$ as an approximation of $g_h$. By using the exact formula of Fourier coefficients for the multivariate biweight kernel $K$, we could efficiently approximate $g_h$. A similar approach can be applied to evaluate remaining functions. Although this approximation theoretically makes sense, we cannot help but using a finite value for $L$ so that the approximation error remains. Another reason is that we should plug the estimated parameters in the limiting variance while the true limiting variance should be calculated based on the true tail parameters. We should not also overlook the fact that there would always be difference between the limiting distribution and the finite sample distribution. Nevertheless, we can still consider to use this approximated statistical inference approach as a reference.

Next experiment is conducted based on the anisotropic Matérn covariance structure. Recall that the covariance function is given as

$$C(\mathbf{s}; \sigma^2, \nu, A, a) = \frac{\sigma^2 (a|A\mathbf{s}|)^\nu}{\Gamma(\nu) 2^{\nu-1}} \mathcal{K}_\nu(a|A\mathbf{s}|)$$

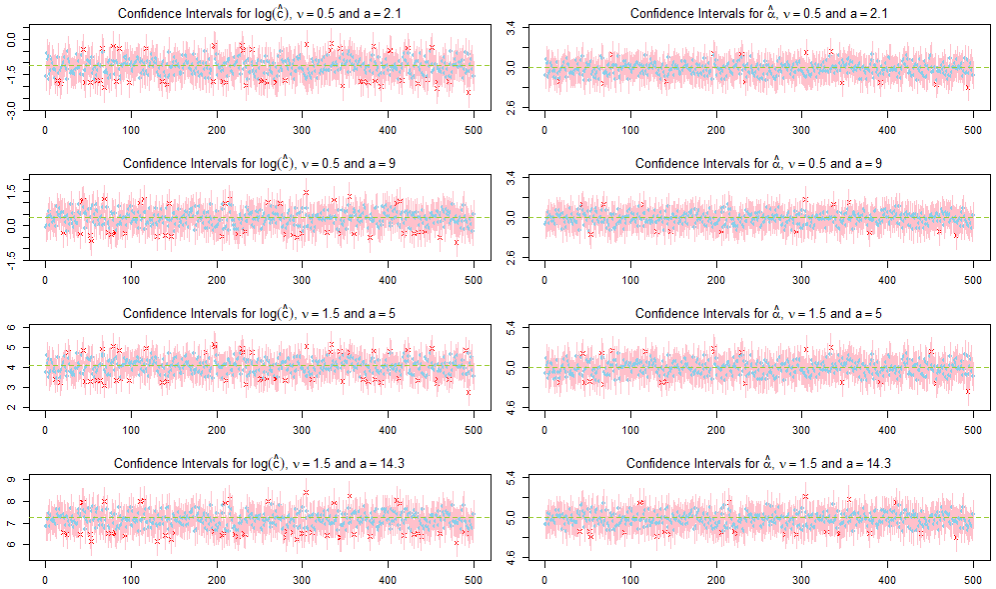


Figure 7.4 Approximated 95% confidence intervals for estimated $\log c$ and $\alpha$. For each of 500 replication, estimated value is depicted by either skyblue dot or red cross. Skyblue dot means the estimated confidence interval actually contains the true value, while red cross means the opposite. Horizontal green dotted lines are drawn at corresponding true values.

so that the spectral density satisfies

$$f(\boldsymbol{\omega};\sigma^2,\nu,A,a) \sim \frac{\sigma^2 a^{2\nu}\Gamma(\nu+\frac{d}{2})}{\pi^{\frac{d}{2}}\Gamma(\nu)}|A^{-T}\boldsymbol{\omega}|^{-2\nu-d} = c|A^{-T}\boldsymbol{\omega}|^{-\alpha}.$$

For identifiable parametrization, $A$ is restricted to upper triangular matrices whose diagonals are positive and determinant is 1 (Anderes, 2010). Sampling scheme is the same as before with $N = 100$ and now two values are considered for $\tau$ which is 2 and 3. Tapered periodograms are used to construct $L_N$ with $M = 10$. Simulations are made based on $\sigma = 1.5, a = 0.8, \nu = 1.75, A_{11} = 1.2$

| $\nu$ | $a$ | Coverage Prob. | |
|---|---|---|---|
| | | $\log c$ | $\alpha$ |
| 0.5 | 2.1 | 0.890 | 0.974 |
| | 9 | 0.910 | 0.972 |
| 1.5 | 5 | 0.890 | 0.958 |
| | 14.3 | 0.892 | 0.952 |

Table 7.2 Coverage probabilities calculated via approximated 95% confidence intervals.

and $A_{12} = 0.5$, which are the same setting as in the simulation study of Anderes (2010). With 500 independent realizations, we first compared our periodogram based method to the squared increment method proposed in Anderes (2010) assuming $\nu$ is known.

| | $\widehat{\sigma^2 a^{2\nu}}$ | | | $\hat{A}_{11}$ | | | $\hat{A}_{12}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | bias | st. dev. | RMSE | bias | st. dev. | RMSE | bias | st. dev. | RMSE |
| Tail Est., $\tau = 2$ | -0.00010 | 0.01660 | 0.01658 | -0.00048 | 0.00612 | 0.00613 | -0.00008 | 0.00873 | 0.00872 |
| Anderes, $m = 2$ | -0.10211 | 0.06596 | 0.12152 | -0.00329 | 0.02423 | 0.02442 | 0.00005 | 0.03421 | 0.03418 |
| Tail Est., $\tau = 3$ | -0.00255 | 0.02214 | 0.02226 | 0.00055 | 0.00833 | 0.00834 | 0.00041 | 0.01179 | 0.01178 |
| Anderes, $m = 3$ | -0.00200 | 0.01542 | 0.01553 | 0.00031 | 0.00561 | 0.00562 | 0.00013 | 0.00811 | 0.00811 |

Table 7.3 Biases, standard deviations and root mean squared errors for parameter estimates under the known $\nu$.

Through the histograms in Figure 7.5, we can observe that using 2 increments in the squared increment method seems to produce biased estimates for the microergodic parameter $\sigma^2 a^{2\nu}$ as well as estimates with larger variances for $A_{11}$ and $A_{12}$ compared to the others. This result coincides with the pre-

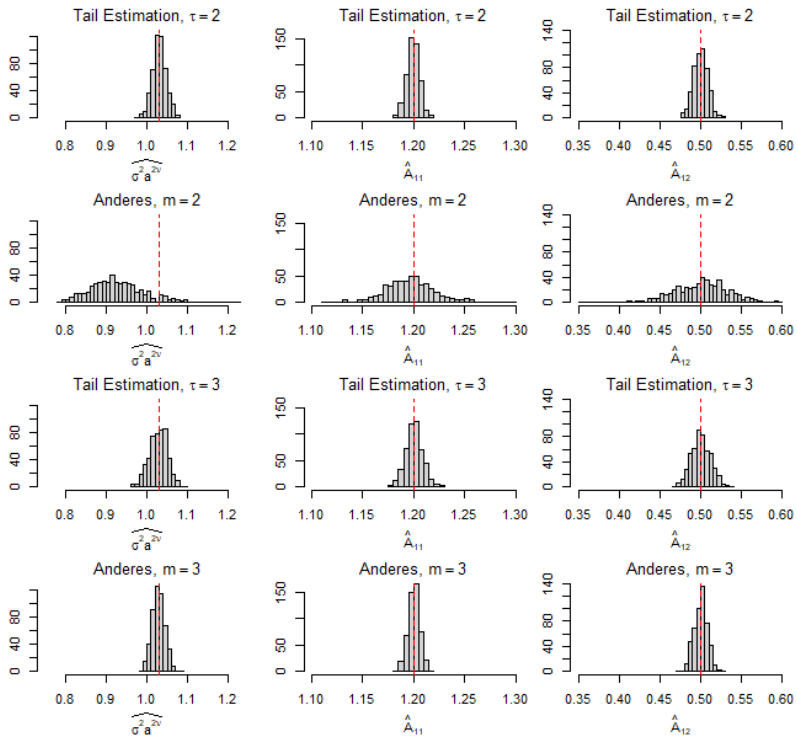


Figure 7.5 Estimation result of $\sigma^2 a^{2\nu}$, $A_{11}$ and $A_{12}$ assuming $\nu = 1.75$ is known. Red vertical lines indicate the locations of true values.

viously reported result in Anderes (2010). In the periodogram based method, both $\tau = 2$ and $\tau = 3$ show satisfactory results, yet $\tau = 2$ seems a slightly better performance in its variance than $\tau = 3$. We also present quantitative comparisons in Table 7.3.

Using the same samples with $\tau = 2$, we apply our method to estimate the smoothness parameter together and display the result in Figure 7.6 and Table 7.4. Note that now $c = \frac{\sigma^2 a^{2\nu} \Gamma(\nu + \frac{d}{2})}{\pi^{\frac{d}{2}} \Gamma(\nu)}$ so that its meaning is somewhat different from the traditional microergodic parameter $\sigma^2 a^{2\nu}$. We can see that all the tail

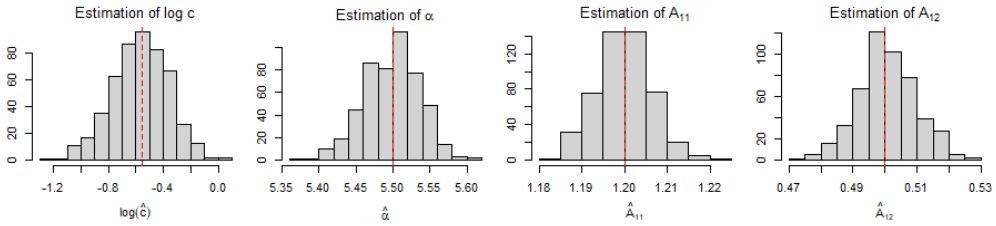


Figure 7.6 Tail parameter estimates of $\log c, \alpha, A_{11}$ and $A_{12}$ using periodogram based method. Red vertical lines indicate the locations of true values.

| | bias | st. dev. | RMSE |
|---|---|---|---|
| $\log c$ | -0.01152 | 0.20683 | 0.20695 |
| $\alpha$ | -0.00187 | 0.03746 | 0.03747 |
| $A_{11}$ | 0.00011 | 0.00634 | 0.00633 |
| $A_{12}$ | 0.00064 | 0.00900 | 0.00900 |

Table 7.4 Biases, standard deviations and root mean squared errors of estimated tail parameters.

parameters are estimated well.

# Chapter 8

# Real Data Analysis

## 8.1 Sea Ice Profiles Data

The first data example is an application to smoothness estimation of Arctic sea ice profiles which had been investigated in Gneiting et al. (2012) and Wu and Lim (2016). Available from `https://nsidc.org/data/g01360`, the data consists of upward looking sonar sea ice draft data collected by submarines. To compare with the results in the literature, we choose the same six profiles each of whose length is 1024 as displayed in Figure 8.1. In spite of the clear non-Gaussianity and/or non-stationarity, the previous analyses do not apply any preprocessing such as transformation. Hence for fair comparison, we also applied our periodogram method directly to the raw data.

Gneiting et al. (2012) and Wu and Lim (2016) estimated smoothness of these data through several methods and presented the result in the form of *fractal*

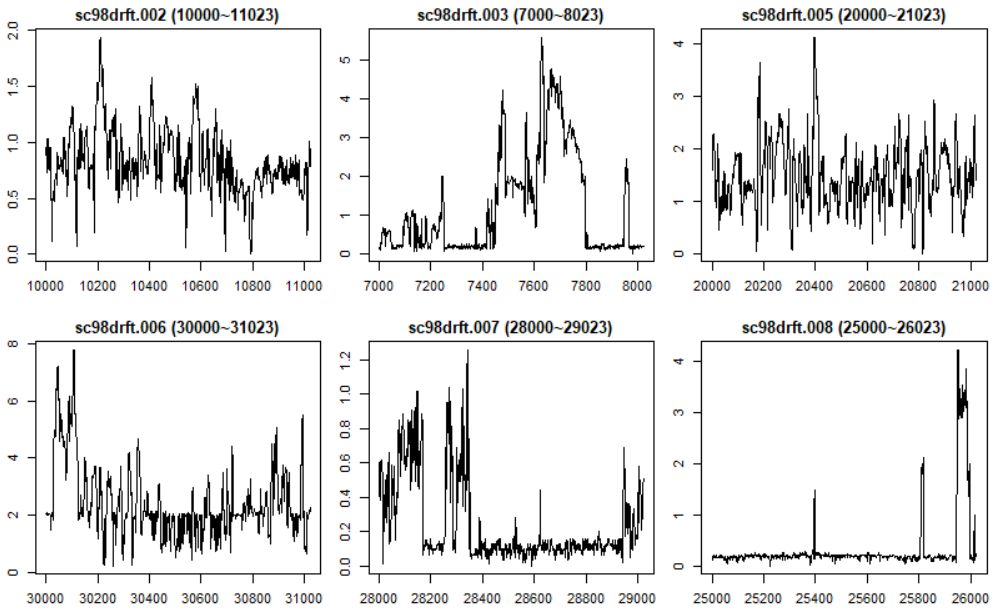


Figure 8.1 Six profiles of Arctic sea ice. Each is comprised of 1024 observations and the distance between two neighboring observations is 1 meter.

*dimension*. The fractal or Hausdorff dimension is a notion that measures roughness of geometric objects. When a function on $\mathbb{R}^d$ which takes values from $\mathbb{R}$ is smooth and differentiable, its graph embedded in $\mathbb{R}^{d+1}$ has the fractal dimension $d$ which is the same as its topological dimension. When such a function is nondifferentiable and rough, however, its fractal dimension takes values between $d$ and $d+1$ which is definitely non-integer. The rougher the graph is, the larger the fractal dimension becomes.

As in Wu and Lim (2016), we assume $f(\omega) \sim c|\omega|^{-\alpha}$ with $\alpha > d = 1$ to each profile. When $\alpha \in (1, 3]$, the fractal dimension of sample paths whose spectral density satisfies the tail condition becomes $D = \frac{5-\alpha}{2} \in [1, 2)$ (see Gneiting et al.,

2012, Table 2). Considering that all the 6 profiles look nondifferentiable, we may expect $D \in (1,2)$ which means $\alpha < 3$. This again implies that $4\tau - \alpha > 1$ for $\tau = 1$ which is enough for $g(\cdot; c, \alpha)$ to be in $C^1(\mathbb{T})$ where

$$g(\omega; c, \alpha) = c\left\{4\sin^2\left(\frac{\omega}{2}\right)\right\}^{2\tau} \sum_{Q\in\mathbb{Z}} |\omega + 2\pi Q|^{-\alpha}.$$

Hence $\tau = 1$ is acceptable choice to apply our method. Since the neighboring observations are 1 meter apart, we take $\delta = 0.001$km. Note that this choice is not important, because the scale of $\delta$ only changes the estimate of $c$ and does not affect the estimate of $\alpha$, and our interest in this analysis is focused on the estimation of $\alpha$. Each sample contains 1024 observations, while actual $N$ gets smaller after applying discrete Laplacian operator so that $N = 1022$ when $\tau = 1$. Tapered periodograms are used to construct $L_N$ with $M = 25, 50, 100, 200$ and $t < \frac{2\pi}{200}$ so that $\Omega_M = \left\{\frac{2\pi J}{M} : J = 1, 2, \cdots, M-1\right\}$. The complete result for this can be found in Table C.1 in the appendix, while we only present here the average of four estimates as our final estimates in the form of fractal dimension in Table 8.1. We also compare the result with those in Gneiting et al. (2012) and Wu and Lim (2016) in Table 8.1, from which we can see that the estimated fractal dimensions from our method mostly concur with the previous results.

## 8.2 Monthly Maximum Temperature Data

For the second data example, we used TerraClimate data which is comprised of worldwide monthly climate information. Using climatically aided interpolation and combining high-spatial resolution with coarser temporal resolution data, TerraClimate produces monthly climate dataset for global terrestrial surfaces

| Method | sc98drft.002 | sc98drft.003 | sc98drft.005 | sc98drft.006 | sc98drft.007 | sc98drft.008 |
|---|---|---|---|---|---|---|
| Variogram | 1.43 | 1.37 | 1.38 | 1.40 | 1.42 | 1.30 |
| Variation(1) | 1.37 | 1.38 | 1.29 | 1.32 | 1.42 | 1.33 |
| Variation(0.5) | 1.32 | 1.37 | 1.24 | 1.27 | 1.38 | 1.31 |
| Hall-Wood | 1.39 | 1.35 | 1.30 | 1.30 | 1.43 | 1.32 |
| Wu & Lim (2016) | 1.47 | 1.32 | 1.32 | 1.39 | 1.41 | 1.31 |
| Tapered Periodogram | 1.38 | 1.30 | 1.31 | 1.36 | 1.38 | 1.18 |

Table 8.1 Comparison of estimated fractal dimensions between various methods. The last row is from our proposed method while the rest are adopted from Wu and Lim (2016).

with spatial resolution of $\frac{1}{24}^\circ$. The data are available from `https://www.climatologylab.org/terraclimate.html` and Abatzoglou et al. (2018) can be referred for further details. We analyze monthly maximum temperature in January 2021. Among the global terrestrial surfaces, relatively local areas are selected by limiting longitude from -120 to -95 and latitude from 35 to 60. The grid points included in this area is of size 600 by 600. We additionally take 2 more layers of grid points surrounding this selected area to maintain $N = 600$ after discrete Laplacian operator is applied with $\tau \leq 2$. A heatmap for worldwide data distribution and the selected area are in Figure 8.2.

We first consider a stationary model for the data whose spectral density has tail behavior $f(\boldsymbol{\omega}) \sim c|A^{-T}\boldsymbol{\omega}|^{-\alpha}$ where $\alpha > d = 2$ and $A$ is an upper triangular matrix with positive diagonals and determinant 1. Using the smoothed periodogram with a multivariate biweight kernel and $N = 600, \tau = 2, \delta = \frac{1}{24}, h = \frac{\pi}{M}$ with $M = 25, 50, 100, 200$, we estimated tail parameters which are summarized in Table 8.2. For $t > 0$ involved in the construction of $\Omega_M$, we choose $t = 0.5$.

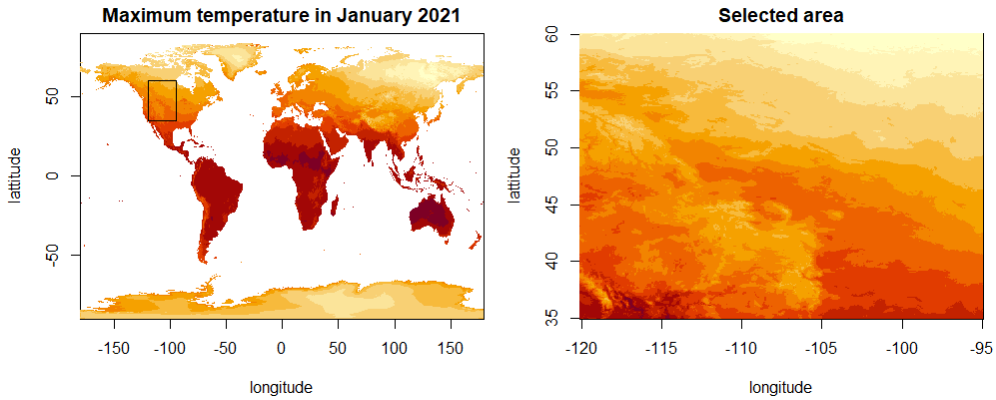


Figure 8.2 The left is the heatmap of the worldwide maximum temperature on global terrestrial surfaces in January 2021. The red corresponds to higher temperature while the yellow to lower temperature. Box with the black boundary line is the selected area of interest which is magnified on the right. Since we add buffers on the boundary of the selected area, the data we are working with actually contains $604 \times 604$ temperature values.

Although we do not present the result here, we also performed the estimation under $\tau = 1$ and achieved the similar estimated result. However theoretical backgrounds we established in the former chapters are only valid when $4\tau - \alpha > 0$, which is violated when $\tau = 1$ and satisfied when $\tau = 2$ since the estimated $\alpha$ were always located between 4 and 5. This is the reason why we choose $\tau = 2$. We also calculated approximate 95% confidence intervals for each parameter. It seems that different choices of $M$ return similar estimates for all the parameters and approximate confidence intervals are mostly overlapped. Moreover all the estimated confidence intervals for $A_{11}$ and $A_{12}$ do not contain 1 and 0 respectively, which means that the geometric anisotropy in the data

is statistically significant. As in the example of sea ice profile data, we may consider the naive average of the four estimates as our final estimates which results in $(\log \hat{c}, \hat{\alpha}, \hat{A}_{11}, \hat{A}_{12}) = (5.6874, 4.8491, 0.9316, 0.0364)$.

| | $\log c$ | $\alpha$ | $A_{11}$ | $A_{12}$ |
|---|---|---|---|---|
| $M = 25$ | 5.6115 | 4.8313 | 0.9286 | 0.0365 |
| | (5.5084, 5.7145) | (4.8152, 4.8474) | (0.9263, 0.9309) | (0.0310, 0.0420) |
| $M = 50$ | 5.6733 | 4.8452 | 0.9337 | 0.0349 |
| | (5.5750, 5.7717) | (4.8299, 4.8606) | (0.9315, 0.9359) | (0.0295, 0.0404) |
| $M = 100$ | 5.7307 | 4.8595 | 0.9319 | 0.0374 |
| | (5.6330, 5.8283) | (4.8443, 4.8748) | (0.9298, 0.9340) | (0.0320, 0.0429) |
| $M = 200$ | 5.7342 | 4.8605 | 0.9324 | 0.0368 |
| | (5.6367, 5.8317) | (4.8452, 4.8757) | (0.9303, 0.9345) | (0.0314, 0.0423) |

Table 8.2 Tail parameter estimates for the monthly maximum temperature in January 2021 over the selected area. Confidence intervals are denoted with parentheses under the corresponding estimates.

We can also consider to apply the method locally as if the data is locally stationary. Setting a local window on the data, we apply the method locally by sliding the window. The size of the window is $5^\circ \times 5^\circ$ so that it consists of $120 \times 120$ grid points, and we slide the window by $2.5^\circ$ along each direction so that entirely $9 \times 9$ different local blocks are obtained. We choose $\tau = 2$ based on the previous estimation result and consider to include additional buffers surrounding the window, which makes each block contain $124 \times 124$ data. After applying the discrete Laplacian operator, we get $N = 120$. As same as before, $\delta$ is set as $\frac{1}{24}$. Also in this analysis, we make use of the smoothed periodogram

with a multivariate biweight kernel and a bandwidth $h = \frac{\pi}{M}$ where $M = 25$. The number $t > 0$ involved in the construction of $\Omega_M$ is chosen as 0.5 as before. The estimated values on each block can be found at tables from Table C.2 to Table C.5 in the appendix. We further calculate confidence intervals on each block and investigate whether spatial variation observed in the estimated parameters is statistically significant. Approximate 95% confidence intervals for each blocks are drawn in Figure 8.3. The result implies that the spectral traits of monthly maximum temperature in January 2021 on the selected area are spatially varying.

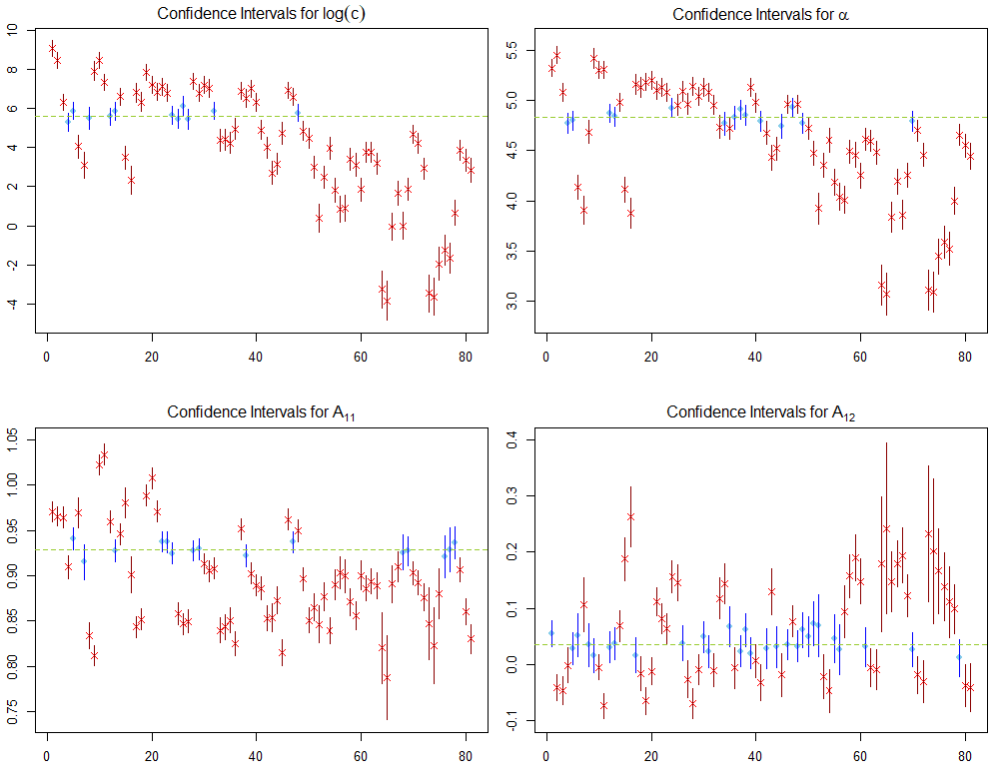


Figure 8.3 Blockwise tail parameter estimates and their approximated 95% confidence intervals. Blue dots and red crosses are estimated values on each block and vertical bars are corresponding confidence intervals. Assuming that the true parameters are $(\log c, \alpha, A_{11}, A_{12}) = (5.6874, 4.8491, 0.9316, 0.0364)$ which are depicted by green dotted horizontal lines, we draw with blue color when the corresponding confidence interval contains the true value while with red color when it does not.

# Chapter 9

# Conclusion and Discussion

In this paper, we developed a method to estimate spectral tail parameters in stationary Gaussian random fields under fixed domain asymptotics with theoretical justifications. The method works on the spectral domain, which is different from the most preexisting methods in fixed domain asymptotic literature to the best of our knowledge. Moreover, our method only assume tail behavior of the spectral density so that it is not restricted to a specific parametric model. Even under the parametric model assumption, our method is meaningful in that it provides a solution to estimate a smoothness parameter and a microergodic parameter simultaneously and consistently which have been rarely investigated. An exception would be Loh et al. (2021), however it only deals with the isotropic Matérn class while ours can also be applied to other classes such as the anisotropic Matérn class or the generalized Wendland class. Last but not least, we derived the asymptotic order and normality of our estimator.

We should admit that there are several limitations in our approach. The one is that all the theories and methods are valid only for complete lattice data. It is currently vague how to generalize our periodogram theories and estimation method for irregular data. Motivated from several researches such as Fuentes (2007), Matsuda and Yajima (2009) and Rao (2018) which investigated periodograms for irregularly spaced data, we may also seek a way to extend our theoretical results and the method for irregular data. Another is that the order of Fejér kernel $M$ or the bandwidth of a compactly supported kernel $h$ appears in the periodogram smoothing stage is considered to be fixed without any criterion on how to choose them. Investigations into how to choose them and what happens when it can vary depending on $N$ can be another direction to work on. Finally, as explained in Chapter 4, we should also admit that we used a locally smoothed version of Whittle likelihood $L_N(\boldsymbol{\theta})$ instead of $\mathcal{L}_N(\boldsymbol{\theta})$ because it was not able to prove the validity of $\mathcal{L}_N(\boldsymbol{\theta})$ from Theorem 4.1. Even so, we could observe from the simulation experiments not shown here that a discretized version of $\mathcal{L}_N$ given as

$$\sum_{\boldsymbol{\omega}\in\Omega_N}\left\{\frac{\mathcal{I}_{N,\delta}(\boldsymbol{\omega})}{\delta^{\alpha-d}g(\boldsymbol{\omega};\boldsymbol{\theta})}+\log\left(\delta^{\alpha-d}g(\boldsymbol{\omega};\boldsymbol{\theta})\right)\right\}$$

where $\Omega_N=\{\frac{2\pi\mathbf{J}}{N}:\mathbf{J}\in\mathbb{Z}_N^d\}\setminus(-t,t)^d$ for some small $t>0$ also works well. This version is much more convenient to implement than the current method, but its theoretical validation would require extremely complicated calculations so that we cannot currently guarantee that it is possible. Nonetheless, if it can be done, it would be a worthwhile contribution for spectral analysis under fixed domain asymptotics.

As a final remark, we would like to mention a couple of future directions that our work can be extended. One is to use different objective functions other than Whittle-type likelihood on the tail parameter estimation stage. Since our estimation method is based on the consistency of $\delta^{d-\alpha_0}\hat{\mathcal{I}}_{N,\delta,M}$ to $g_{N,M}(\cdot;\boldsymbol{\theta}_0)$ and the main idea for estimation is to make the ratio of $\hat{\mathcal{I}}_{N,\delta,M}$ to $\delta^{\alpha-d}g_{N,M}(\cdot;\boldsymbol{\theta})$ as close to 1 as possible, we can conjecture that any objective function that operates in a similar way would work well. For example, we may consider so called least square method which minimizes

$$\sum_{\boldsymbol{\omega}\in\Omega_M}\left(\frac{\hat{\mathcal{I}}_{N,\delta,M}(\boldsymbol{\omega})}{\delta^{\alpha-d}g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})}-1\right)^2.$$

The properties of the estimator will of course differ by which objective function is used. Various investigations focusing on different perspectives, such as on which objective function the estimator achieves asymptotically minimal variance or becomes robust under contamination by measurement error or outliers, would be additional possible directions.

Another is a multivariate extension. Although all the equations and notations will become more complicated and tedious, the extension of our smoothed periodogram will be straightforward under mild conditions on cross spectral densities motivated from Lim and Stein (2008). In this way, we expect to get the consistency result for a smoothed periodogram matrix $\hat{\boldsymbol{\mathcal{I}}}_{N,\delta,M}$ after appropriate scaling involving $\delta$ and decaying rate of all the spectral densities and cross spectral densities. For the tail parameter estimation, we can adopt the idea of Whittle (1953) which involves the trace and determinant of matrix-valued functions. Since there are few theoretical discoveries related to multivariate random

fields under fixed domain asymptotics, this direction of extension would serve a new contribution in the area.

The last is about the separate estimability of variance and scale parameters in the Matérn covariance class when $d \geq 5$ which had been investigated in Anderes (2010). Although the scale parameter, $a$, is not a tail parameter, we can use the approach in Chapter 6 to estimate $\sigma^2$ and $a$ separately. From the sketch not included in this work, it can be derived for the isotropic Matérn class that under some mild conditions, minimizing $R_N(c, \alpha, a)$ with respect to all three parameters gives consistent estimates $(\hat{c}, \hat{\alpha}, \hat{a})$ with orders $\hat{c} - c_0 = O_P(N^{-\frac{d}{2}} \log N), \hat{\alpha} - \alpha_0 = O_P(N^{-\frac{d}{2}})$ and $\hat{a} - a_0 = O_P(N^{-\frac{d}{2}+2})$ when $d \geq 5$. Rigorous organization about this informal statement as well as including geometric anisotropy as in Anderes (2010) can be a straightforward but complicated extension of this work.

# Appendix A

# Proofs for Chapter 4

To show the limiting distribution becomes Gaussian in Theorem 4.1, we will show that the second order cumulants converge to some limit values and all the cumulants of order 3 and higher converge to zero. For this end, we first evaluate cumulants. For $p$ smoothed periodograms

$$\int_{\mathbb{T}^d} \phi_1(\boldsymbol{\omega}) \mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega}) \mathbf{d}\boldsymbol{\omega}, \cdots, \int_{\mathbb{T}^d} \phi_p(\boldsymbol{\omega}) \mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega}) \mathbf{d}\boldsymbol{\omega},$$

their $m$th order joint cumulant can be any of the form

$$\operatorname{cum}\left( \int_{\mathbb{T}^d} \phi_{k_1}(\boldsymbol{\omega}) \mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega}) \mathbf{d}\boldsymbol{\omega}, \cdots, \int_{\mathbb{T}^d} \phi_{k_m}(\boldsymbol{\omega}) \mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega}) \mathbf{d}\boldsymbol{\omega} \right)$$

for $k_1, \cdots, k_m \in \{1, \cdots, p\}$. The following lemmas are helpful.

**Lemma A.1.**

$$\begin{aligned}
&\mathrm{cum}\left(\int_{\mathbb{T}^d}\phi_{k_1}(\boldsymbol{\omega})\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega},\cdots,\int_{\mathbb{T}^d}\phi_{k_m}(\boldsymbol{\omega})\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega}\right)\\
&=\sum_{(\boldsymbol{\iota},\mathbf{e})}\int_{\mathbb{T}^{2md}}\phi_{k_1}(\boldsymbol{\omega}_1)\phi_{k_2}((-1)^{\bar{e}_2}\boldsymbol{\omega}_{\iota_2})\cdots\phi_{k_m}((-1)^{\bar{e}_m}\boldsymbol{\omega}_{\iota_m})\bar{f}^{\tau}_{\delta}(\boldsymbol{\upsilon}_1)\cdots\bar{f}^{\tau}_{\delta}(\boldsymbol{\upsilon}_m)\\
&\qquad\times(2\pi N)^{-md}P_N(\boldsymbol{\omega}_1,\boldsymbol{\upsilon}_1,\boldsymbol{\omega}_2,\boldsymbol{\upsilon}_2,\cdots,\boldsymbol{\omega}_m,\boldsymbol{\upsilon}_m)\mathbf{d}\boldsymbol{\omega}_1\cdots\mathbf{d}\boldsymbol{\omega}_m\mathbf{d}\boldsymbol{\upsilon}_1\cdots\mathbf{d}\boldsymbol{\upsilon}_m,
\end{aligned}$$

*where* $(\boldsymbol{\iota},\mathbf{e})=(\iota_1,\cdots,\iota_m,e_1,\cdots,e_m)$ *such that* $(\iota_1,\cdots,\iota_m)$ *is a permutation of a number from 1 to* $m$ *with* $\iota_1=1$ *and* $\{e_k,\bar{e}_k\}=\{1,2\},k=1,\cdots,m$ *with* $e_1=1$ *and*

$$\begin{aligned}
&P_N(\boldsymbol{\omega}_1,\boldsymbol{\upsilon}_1,\boldsymbol{\omega}_2,\boldsymbol{\upsilon}_2,\cdots,\boldsymbol{\omega}_m,\boldsymbol{\upsilon}_m)\\
&=\sum_{\mathbf{J}_{11},\cdots,\mathbf{J}_{m2}\in\mathbb{Z}^d_N}e^{-i\langle\boldsymbol{\omega}_1-\boldsymbol{\upsilon}_1,\mathbf{J}_{11}\rangle}e^{-i\langle\boldsymbol{\upsilon}_1-\boldsymbol{\omega}_2,\mathbf{J}_{22}\rangle}e^{-i\langle\boldsymbol{\omega}_2-\boldsymbol{\upsilon}_2,\mathbf{J}_{21}\rangle}e^{-i\langle\boldsymbol{\upsilon}_2-\boldsymbol{\omega}_3,\mathbf{J}_{32}\rangle}\\
&\qquad\qquad\times\cdots\times e^{-i\langle\boldsymbol{\omega}_m-\boldsymbol{\upsilon}_m,\mathbf{J}_{m1}\rangle}e^{-i\langle\boldsymbol{\upsilon}_m-\boldsymbol{\omega}_1,\mathbf{J}_{12}\rangle}.
\end{aligned}$$

*Proof.* From

$$\int_{\mathbb{T}^d}\phi(\boldsymbol{\omega})\mathcal{I}^{\tau}_{N,\delta}(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega}=(2\pi N)^{-d}\sum_{\mathbf{J},\mathbf{K}\in\mathbb{Z}^d_N}Y^{\tau}_{\delta}(\mathbf{J})Y^{\tau}_{\delta}(\mathbf{K})\int_{\mathbb{T}^d}\phi(\boldsymbol{\omega})e^{-i\langle\boldsymbol{\omega},\mathbf{J}-\mathbf{K}\rangle}\mathbf{d}\boldsymbol{\omega},$$

the cumulant we are interested in is represented as $(2\pi N)^{-md}$ times

$$\begin{aligned}
&\sum_{\mathbf{J}_{11},\cdots,\mathbf{J}_{m2}\in\mathbb{Z}^d_N}\mathrm{cum}\left(Y^{\tau}_{\delta}(\mathbf{J}_{11})Y^{\tau}_{\delta}(\mathbf{J}_{12}),\cdots,Y^{\tau}_{\delta}(\mathbf{J}_{m1})Y^{\tau}_{\delta}(\mathbf{J}_{m2})\right)\\
&\times\int_{\mathbb{T}^{md}}\phi_{k_1}(\boldsymbol{\omega}_1)\cdots\phi_{k_m}(\boldsymbol{\omega}_m)e^{-i\langle\boldsymbol{\omega}_1,\mathbf{J}_{11}-\mathbf{J}_{12}\rangle}\cdots e^{-i\langle\boldsymbol{\omega}_m,\mathbf{J}_{m1}-\mathbf{J}_{m2}\rangle}\mathbf{d}\boldsymbol{\omega}_1\cdots\mathbf{d}\boldsymbol{\omega}_m.
\end{aligned}$$

The cumulant term inside the sum is equal to

$$\begin{aligned}
&\operatorname{cum}\left(Y_\delta^\tau(\mathbf{J}_{11})Y_\delta^\tau(\mathbf{J}_{12}),\cdots,Y_\delta^\tau(\mathbf{J}_{m1})Y_\delta^\tau(\mathbf{J}_{m2})\right)\\
&=\sum_{(\boldsymbol{\iota},\mathbf{e})}\operatorname{cum}\left\{Y_\delta^\tau(\mathbf{J}_{\iota_1 e_1}),Y_\delta^\tau(\mathbf{J}_{\iota_2\bar{e}_2})\right\}\times\operatorname{cum}\left\{Y_\delta^\tau(\mathbf{J}_{\iota_2 e_2}),Y_\delta^\tau(\mathbf{J}_{\iota_3\bar{e}_3})\right\}\times\cdots\\
&\qquad\times\operatorname{cum}\left\{Y_\delta^\tau(\mathbf{J}_{\iota_{m-1}e_{m-1}}),Y_\delta^\tau(\mathbf{J}_{\iota_m\bar{e}_m})\right\}\times\operatorname{cum}\left\{Y_\delta^\tau(\mathbf{J}_{\iota_m e_m}),Y_\delta^\tau(\mathbf{J}_{\iota_1\bar{e}_1})\right\}
\end{aligned}$$

where the sum runs over $(\boldsymbol{\iota},\mathbf{e})=(\iota_1,\cdots,\iota_m,e_1,\cdots,e_m)$ such that $(\iota_1,\cdots,\iota_m)$ is a permutation of a number from 1 to $m$ with $\iota_1=1$ and $\{e_k,\bar{e}_k\}=\{1,2\}, k=1,\cdots,m$ with $e_1=1$. (See Theorem 7 of Lim and Stein (2008) or Theorem 2.3.2 of Brillinger (2001).) Since

$$\operatorname{cum}\left\{Y_\delta^\tau(\mathbf{J}_{\iota_k e_k}),Y_\delta^\tau(\mathbf{J}_{\iota_{k+1}\bar{e}_{k+1}})\right\}=\int_{\mathbb{T}^d}\bar{f}_\delta^\tau(\boldsymbol{\upsilon})e^{i\langle\boldsymbol{\upsilon},\mathbf{J}_{\iota_k e_k}-\mathbf{J}_{\iota_{k+1}\bar{e}_{k+1}}\rangle}\mathbf{d}\boldsymbol{\upsilon},$$

the entire cumulant is now $(2\pi N)^{-md}$ times

$$\begin{aligned}
&\sum_{(\boldsymbol{\iota},\mathbf{e})}\int_{\mathbb{T}^{2md}}\phi_{k_1}(\boldsymbol{\omega}_1)\cdots\phi_{k_m}(\boldsymbol{\omega}_m)\bar{f}_\delta^\tau(\boldsymbol{\upsilon}_{\iota_1})\cdots\bar{f}_\delta^\tau(\boldsymbol{\upsilon}_{\iota_m})\\
&\quad\times\sum_{\mathbf{J}_{11},\cdots,\mathbf{J}_{m2}\in\mathbb{Z}_N^d}e^{-i\langle\boldsymbol{\omega}_1,\mathbf{J}_{11}-\mathbf{J}_{12}\rangle-i\langle\boldsymbol{\omega}_2,\mathbf{J}_{21}-\mathbf{J}_{22}\rangle-\cdots-i\langle\boldsymbol{\omega}^m,\mathbf{J}_{m1}-\mathbf{J}_{m2}\rangle}\\
&\qquad\qquad\times e^{i\langle\boldsymbol{\upsilon}_{\iota_1},\mathbf{J}_{\iota_1 e_1}-\mathbf{J}_{\iota_2\bar{e}_2}\rangle+i\langle\boldsymbol{\upsilon}_{\iota_2},\mathbf{J}_{\iota_2 e_2}-\mathbf{J}_{\iota_3\bar{e}_3}\rangle+\cdots+i\langle\boldsymbol{\upsilon}_{\iota_m},\mathbf{J}_{\iota_m e_m}-\mathbf{J}_{\iota_1\bar{e}_1}\rangle}\\
&\quad\times\mathbf{d}\boldsymbol{\omega}_1\cdots\mathbf{d}\boldsymbol{\omega}_m\mathbf{d}\boldsymbol{\upsilon}_1\cdots\mathbf{d}\boldsymbol{\upsilon}_m.
\end{aligned}$$

This can be rewritten as

$$\begin{aligned}
&\sum_{(\boldsymbol{\iota},\mathbf{e})}\int_{\mathbb{T}^{2md}}\phi_{k_1}(\boldsymbol{\omega}_1)\cdots\phi_{k_m}(\boldsymbol{\omega}_m)\bar{f}_\delta^\tau(\boldsymbol{\upsilon}_{\iota_1})\cdots\bar{f}_\delta^\tau(\boldsymbol{\upsilon}_{\iota_m})\\
&\quad\times\sum_{\mathbf{J}_{11},\cdots,\mathbf{J}_{m2}\in\mathbb{Z}_N^d} e^{i\langle(-1)^{\bar{e}_1}\boldsymbol{\omega}_{\iota_1},\mathbf{J}_{\iota_1\bar{e}_1}-\mathbf{J}_{\iota_1 e_1}\rangle+\cdots+i\langle(-1)^{\bar{e}_p}\boldsymbol{\omega}_{\iota_m},\mathbf{J}_{\iota_m\bar{e}_m}-\mathbf{J}_{\iota_m e_m}\rangle}\\
&\qquad\qquad\times e^{i\langle\boldsymbol{\upsilon}_{\iota_1},\mathbf{J}_{\iota_1 e_1}-\mathbf{J}_{\iota_2\bar{e}_2}\rangle+i\langle\boldsymbol{\upsilon}_{\iota_2},\mathbf{J}_{\iota_2 e_2}-\mathbf{J}_{\iota_3\bar{e}_3}\rangle+\cdots+i\langle\boldsymbol{\upsilon}_{\iota_m},\mathbf{J}_{\iota_m e_m}-\mathbf{J}_{\iota_1\bar{e}_1}\rangle}\\
&\quad\times\mathbf{d}\boldsymbol{\omega}_1\cdots\mathbf{d}\boldsymbol{\omega}_m\mathbf{d}\boldsymbol{\upsilon}_1\cdots\mathbf{d}\boldsymbol{\upsilon}_m\\
=&\sum_{(\boldsymbol{\iota},\mathbf{e})}\int_{\mathbb{T}^{2md}}\phi_{k_1}(\boldsymbol{\omega}_1)\cdots\phi_{k_m}(\boldsymbol{\omega}_m)\bar{f}_\delta^\tau(\boldsymbol{\upsilon}_{\iota_1})\cdots\bar{f}_\delta^\tau(\boldsymbol{\upsilon}_{\iota_m})\\
&\quad\times\sum_{\mathbf{J}_{11},\cdots,\mathbf{J}_{m2}\in\mathbb{Z}_N^d} e^{i\langle\boldsymbol{\upsilon}_{\iota_1}-(-1)^{\bar{e}_1}\boldsymbol{\omega}_{\iota_1},\mathbf{J}_{\iota_1 e_1}\rangle}e^{i\langle(-1)^{\bar{e}_2}\boldsymbol{\omega}_{\iota_2}-\boldsymbol{\upsilon}_{\iota_1},\mathbf{J}_{\iota_2\bar{e}_2}\rangle}\\
&\qquad\qquad\times e^{i\langle\boldsymbol{\upsilon}_{\iota_2}-(-1)^{\bar{e}_2}\boldsymbol{\omega}_{\iota_2},\mathbf{J}_{\iota_2 e_2}\rangle}e^{i\langle(-1)^{\bar{e}_3}\boldsymbol{\omega}_{\iota_3}-\boldsymbol{\upsilon}_{\iota_2},\mathbf{J}_{\iota_3\bar{e}_3}\rangle}\\
&\qquad\qquad\times\cdots\\
&\qquad\qquad\times e^{i\langle\boldsymbol{\upsilon}_{\iota_m}-(-1)^{\bar{e}_m}\boldsymbol{\omega}_{\iota_m},\mathbf{J}_{\iota_m e_m}\rangle}e^{i\langle(-1)^{\bar{e}_1}\boldsymbol{\omega}_{\iota_1}-\boldsymbol{\upsilon}_{\iota_m},\mathbf{J}_{\iota_1\bar{e}_1}\rangle}\\
&\quad\times\mathbf{d}\boldsymbol{\omega}_1\cdots\mathbf{d}\boldsymbol{\omega}_m\mathbf{d}\boldsymbol{\upsilon}_1\cdots\mathbf{d}\boldsymbol{\upsilon}_m.
\end{aligned}$$

By change of variables and using that the sum is over $(\boldsymbol{\iota},\mathbf{e})$, the last sum on the above becomes

$$\begin{aligned}
&\sum_{(\boldsymbol{\iota},\mathbf{e})}\int_{\mathbb{T}^{2md}}\phi_{k_1}(\boldsymbol{\omega}_1)\phi_{k_2}((-1)^{\bar{e}_2}\boldsymbol{\omega}_{\iota_2})\cdots\phi_{k_m}((-1)^{\bar{e}_m}\boldsymbol{\omega}_{\iota_m})\bar{f}_\delta^\tau(\boldsymbol{\upsilon}_1)\cdots\bar{f}_\delta^\tau(\boldsymbol{\upsilon}_m)\\
&\times\sum_{\mathbf{J}_{11},\cdots,\mathbf{J}_{m2}\in\mathbb{Z}_N^d} e^{i\langle\boldsymbol{\upsilon}_1-\boldsymbol{\omega}_1,\mathbf{J}_{11}\rangle+i\langle\boldsymbol{\omega}_2-\boldsymbol{\upsilon}_1,\mathbf{J}_{22}\rangle+i\langle\boldsymbol{\upsilon}_2-\boldsymbol{\omega}_2,\mathbf{J}_{21}\rangle+\cdots+i\langle\boldsymbol{\upsilon}_m-\boldsymbol{\omega}_m,\mathbf{J}_{m1}\rangle+i\langle\boldsymbol{\omega}_1-\boldsymbol{\upsilon}_m,\mathbf{J}_{12}\rangle}\\
&\times\mathbf{d}\boldsymbol{\omega}_1\cdots\mathbf{d}\boldsymbol{\omega}_m\mathbf{d}\boldsymbol{\upsilon}_1\cdots\mathbf{d}\boldsymbol{\upsilon}_m.
\end{aligned}$$

□

To calculate the limits of cumulants found in Lemma A.1, we need to investigate what happens on

$$\int_{\mathbb{T}^{sd}} h_1(\mathbf{x}_1)h_2(\mathbf{x}_2)\cdots h_s(\mathbf{x}_s)P_N(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_s)\mathbf{d}\mathbf{x}_1\cdots\mathbf{d}\mathbf{x}_s$$

for some functions $h_1,\cdots,h_s$ when $s=2m$. Note that we have to consider that some of $g_j$ may depend on $N$, since we have $\bar{f}_\delta^\tau$ at the integrand in the cumulant expression. Recall that $P_N$ is of form

$$\begin{aligned}
&P_N(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_s)\\
&=\sum_{\mathbf{J}_1,\cdots,\mathbf{J}_s\in\mathbb{Z}_N^d} e^{-i\langle\mathbf{x}_1-\mathbf{x}_2,\mathbf{J}_1\rangle}e^{-i\langle\mathbf{x}_2-\mathbf{x}_3,\mathbf{J}_2\rangle}\cdots e^{-i\langle\mathbf{x}_{s-1}-\mathbf{x}_s,\mathbf{J}_{s-1}\rangle}e^{-i\langle\mathbf{x}_s-\mathbf{x}_1,\mathbf{J}_s\rangle}\\
&=\prod_{j=1}^{s}\left(\sum_{\mathbf{J}_j\in\mathbb{Z}_N^d} e^{-i\langle\mathbf{x}_j-\mathbf{x}_{j+1},\mathbf{J}_j\rangle}\right)=\prod_{j=1}^{s}\prod_{l=1}^{d}\frac{\sin\left(\frac{N(x_{jl}-x_{(j+1)l})}{2}\right)}{\sin\left(\frac{x_{jl}-x_{(j+1)l}}{2}\right)}
\end{aligned}$$

where $\mathbf{x}_j=(x_{jl})_{l=1,\cdots,d}$ and $s+1$ is identified with 1. Observe that each term in the right side is Dirichlet kernel of order $N$, from which integration with $P_N$ can be thought as $s$-cyclic convolution of $d$-dimensional Dirichlet kernels. Based on this observation, we can think intuitively that

$$\int_{\mathbb{T}^d} h_1(\mathbf{x}_1)P_N(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_s)\mathbf{d}\mathbf{x}_1\approx(2\pi)^d h_1(\mathbf{x}_2)P_N(\mathbf{x}_2,\mathbf{x}_3,\cdots,\mathbf{x}_s)$$

and henceforth

$$\begin{aligned}
&\int_{\mathbb{T}^{2d}} h_1(\mathbf{x}_1)h_2(\mathbf{x}_2)P_N(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_s)\mathbf{d}\mathbf{x}_1\mathbf{d}\mathbf{x}_2\\
&\approx(2\pi)^d\int_{\mathbb{T}^d} h_1(\mathbf{x}_2)h_2(\mathbf{x}_2)P_N(\mathbf{x}_2,\mathbf{x}_3,\cdots,\mathbf{x}_s)\mathbf{d}\mathbf{x}_2\\
&\approx(2\pi)^{2d}h_1(\mathbf{x}_3)h_2(\mathbf{x}_3)P_N(\mathbf{x}_3,\cdots,\mathbf{x}_s).
\end{aligned}$$

After $(s-2)$ repetition, we get

$$\int_{\mathbb{T}^{sd}} h_1(\mathbf{x}_1)\cdots h_s(\mathbf{x}_s)P_N(\mathbf{x}_1,\cdots,\mathbf{x}_s)\mathbf{dx}_1\cdots\mathbf{dx}_s$$
$$\approx (2\pi)^{(s-2)d}\int_{\mathbb{T}^{2d}} h_1(\mathbf{x}_{s-1})\cdots h_{s-1}(\mathbf{x}_{s-1})h_s(\mathbf{x}_s)P_N(\mathbf{x}_{s-1},\mathbf{x}_s)\mathbf{dx}_{s-1}\mathbf{dx}_s.$$

Now, observe that $P_N(\mathbf{x}_{s-1},\mathbf{x}_s)$ is $(2\pi N)^d$ times Fejér kernel of order $N$. Hence we can think of the right hand side similar as

$$(2\pi)^{(s-1)d}N^d\int_{\mathbb{T}^d} h_1(\mathbf{x}_s)\cdots h_s(\mathbf{x}_s)\mathbf{dx}_s.$$

Of course, this intuition will not always be true. Indeed, it is well known that convolution of a function with Dirichlet kernel may fail to converge to the true function value when the function is merely continuous. (See for instance Corollary 1.10 of Muscalu and Schlag (2013).) Under the assumption of continuous differentiability, we show that the above intuitive calculation holds.

**Lemma A.2.** *For $h_1,\cdots,h_s\in C^1(\mathbb{T}^d)$, we have*

$$\left|\frac{1}{(2\pi)^{(s-1)d}N^d}\int_{\mathbb{T}^{sd}} h_1(\mathbf{x}_1)\cdots h_s(\mathbf{x}_s)P_N(\mathbf{x}_1,\cdots,\mathbf{x}_s)\mathbf{dx}_1\cdots\mathbf{dx}_s - \int_{\mathbb{T}^d} h_1(\mathbf{x})\cdots h_s(\mathbf{x})\mathbf{dx}\right|$$
$$\lesssim \sum_{j=1}^{s-1}\left(\|\partial h_j\|_{L^\infty}\prod_{k\neq j}\|h_k\|_{L^\infty}\right)\times\frac{(\log N)^{(s-2)d+1}}{N},$$

*where $\|\partial h\|_{L^\infty}=\sup\left\{\left|\frac{\partial h}{\partial x_l}(\mathbf{x})\right| : l=1,\cdots,d,\mathbf{x}\in\mathbb{T}^d\right\}$. Here $h_1,\cdots,h_s$ may depend on $N$. In that case, the $L^\infty$ norm part on the bound becomes dependent to $N$ as well.*

*Proof.* Since $h_1\in C^1(\mathbb{T}^d)$, we have

$$|h_1(\mathbf{x}_1)-h_1(\mathbf{x}_2)|\leq\|\partial h_1\|_{L^\infty}\sum_{l=1}^{d}|x_{1l}-x_{2l}|.$$

Thus

$$\left|\int_{\mathbb{T}^{pd}} \{h_1(\mathbf{x}_1)h_2(\mathbf{x}_2)\cdots h_s(\mathbf{x}_s) - h_1(\mathbf{x}_2)h_2(\mathbf{x}_2)\cdots h_s(\mathbf{x}_s)\}\, P_N(\mathbf{x}_1,\cdots,\mathbf{x}_s)\mathbf{dx}_1\cdots\mathbf{dx}_s\right|$$
$$\lesssim \|\partial h_1\|_{L^\infty}\prod_{j=2}^{s}\|h_j\|_{L^\infty}\sum_{l=1}^{d}\int_{\mathbb{T}^{sd}}|x_{1l}-x_{2l}||P_N(\mathbf{x}_1,\cdots,\mathbf{x}_s)|\mathbf{dx}_1\cdots\mathbf{dx}_s.$$

Moreover since

$$\frac{\sin(Nx/2)}{\sin(x/2)} \lesssim \frac{1}{N^{-1}+|x|},$$

$P_N$ can be bounded as

$$|P_N(\mathbf{x}_1,\cdots,\mathbf{x}_s)|$$
$$\lesssim \prod_{l=1}^{d}\frac{1}{(N^{-1}+|x_{1l}-x_{2l}|)}\frac{1}{(N^{-1}+|x_{2l}-x_{1l}|)}\cdots\frac{1}{(N^{-1}+|x_{(s-1)l}-x_{sl}|)}\frac{1}{(N^{-1}+|x_{sl}-x_{1l}|)}.$$

Using

$$\int_{\mathbb{T}}\frac{1}{(N^{-1}+|x-y|)}\frac{1}{(N^{-1}+|y-z|)}dy \lesssim \frac{\log N}{N^{-1}+|x-z|}$$

and

$$\int_{\mathbb{T}}\frac{1}{(N^{-1}+|y|)^2}dy \lesssim N,$$

we have

$$\int_{\mathbb{T}^s}\frac{1}{(N^{-1}+|x_{1l}-x_{2l}|)}\frac{1}{(N^{-1}+|x_{2l}-x_{3l}|)}\cdots\frac{1}{(N^{-1}+|x_{(s-1)l}-x_{sl}|)}\frac{1}{(N^{-1}+|x_{sl}-x_{1l}|)}dx_{1l}\cdots dx_{sl}$$
$$\lesssim \int_{\mathbb{T}^{s-1}}\log N\frac{1}{(N^{-1}+|x_{2l}-x_{3l}|)}\cdots\frac{1}{(N^{-1}+|x_{(s-1)l}-x_{sl}|)}\frac{1}{(N^{-1}+|x_{sl}-x_{2l}|)}dx_{2l}\cdots dx_{sl}$$
$$\lesssim \cdots$$
$$\lesssim \int_{\mathbb{T}^2}(\log N)^{s-2}\frac{1}{(N^{-1}+|x_{(s-1)l}-x_{sl}|)}\frac{1}{(N^{-1}+|x_{sl}-x_{(s-1)l}|)}dx_{(s-1)l}dx_{sl}$$
$$\lesssim N(\log N)^{s-2}$$

and

$$\int_{\mathbb{T}^s} \frac{|x_{1l}-x_{2l}|}{(N^{-1}+|x_{1l}-x_{2l}|)} \frac{1}{(N^{-1}+|x_{2l}-x_{3l}|)} \cdots \frac{1}{(N^{-1}+|x_{sl}-x_{1l}|)} dx_{1l}\cdots dx_{sl}$$
$$\leq \int_{\mathbb{T}^s} \frac{1}{(N^{-1}+|x_{2l}-x_{3l}|)} \frac{1}{(N^{-1}+|x_{3l}-x_{4l}|)} \cdots \frac{1}{(N^{-1}+|x_{sl}-x_{1l}|)} dx_{1l}\cdots dx_{sl}$$
$$\lesssim (\log N)^{p-1}.$$

Combining these order estimates, we arrive at

$$\int_{\mathbb{T}^{sd}} |x_{2l}-x_{1l}||P_N(\mathbf{x}_1,\cdots,\mathbf{x}_s)|\mathbf{dx}_1\cdots\mathbf{dx}_s$$
$$\lesssim N^{d-1}(\log N)^{(s-2)(d-1)} \times (\log N)^{s-1} = N^{d-1}(\log N)^{(s-2)d+1}.$$

We can repeat this for the remaining variables to get

$$\left|\int_{\mathbb{T}^{sd}} \{h_1(\mathbf{x}_1)h_2(\mathbf{x}_2)\cdots h_s(\mathbf{x}_s) - h_1(\mathbf{x}_s)h_2(\mathbf{x}_s)\cdots h_s(\mathbf{x}_s)\} P_N(\mathbf{x}_1,\cdots,\mathbf{x}_s)\mathbf{dx}_1\cdots\mathbf{dx}_s\right|$$
$$\lesssim \left( \|\partial h_1\|_{L^\infty} \prod_{j=2}^{s} \|h_j\|_{L^\infty} + \|\partial(h_1h_2)\|_{L^\infty} \prod_{j=3}^{d} \|h_j\|_{L^\infty} + \right.$$
$$\left. \cdots + \|\partial(h_1\cdots h_{s-1})\|_{L^\infty}\|h_s\|_{L^\infty} \right) \times N^{d-1}(\log N)^{(s-2)d+1}$$
$$\lesssim \sum_{j=1}^{s-1} \left( \|\partial h_j\|_{L^\infty} \prod_{k\neq j} \|h_k\|_{L^\infty} \right) \times N^{d-1}(\log N)^{(s-2)d+1}.$$

The last bound is from chain rule. Now, by using another expression for $P_N$, we get

$$\int_{\mathbb{T}^{sd}} h_1(\mathbf{x}_s)h_2(\mathbf{x}_s)\cdots h_s(\mathbf{x}_s)P_N(\mathbf{x}_1,\cdots,\mathbf{x}_s)\mathbf{dx}_1\cdots\mathbf{dx}_s$$
$$= \sum_{\mathbf{J}_1,\cdots,\mathbf{J}_s\in\mathbb{Z}_N^d} \int_{\mathbb{T}^{sd}} h_1(\mathbf{x}_s)h_2(\mathbf{x}_s)\cdots h_s(\mathbf{x}_s)$$
$$\times e^{i\langle \mathbf{x}_1,\mathbf{J}_s-\mathbf{J}_1\rangle} e^{i\langle \mathbf{x}_2,\mathbf{J}_1-\mathbf{J}_2\rangle} \cdots e^{i\langle \mathbf{x}_{s-1},\mathbf{J}_{s-2}-\mathbf{J}_{s-1}\rangle} e^{i\langle \mathbf{x}_s,\mathbf{J}_{s-1}-\mathbf{J}_s\rangle} \mathbf{dx}_1\cdots\mathbf{dx}_s.$$

For each given $\mathbf{J}_1$ and $\mathbf{J}_s$, we can calculate the first $(s-1)$-integrals in the above integration, i.e. with respect to $\mathbf{dx}_1 \cdots \mathbf{dx}_{s-1}$ so that

$$\int_{\mathbb{T}^{(s-1)^d}} e^{i\langle \mathbf{x}_1, \mathbf{J}_s - \mathbf{J}_1\rangle} e^{i\langle \mathbf{x}_2, \mathbf{J}_1 - \mathbf{J}_2\rangle} \cdots e^{i\langle \mathbf{x}_{s-1}, \mathbf{J}_{s-2} - \mathbf{J}_{s-1}\rangle} e^{i\langle \mathbf{x}_s, \mathbf{J}_{s-1} - \mathbf{J}_s\rangle} \mathbf{dx}_1 \cdots \mathbf{dx}_{s-1}$$
$$= \begin{cases} (2\pi)^{(s-1)d} & \text{if } \mathbf{J}_1 = \mathbf{J}_2 = \cdots = \mathbf{J}_{s-1} = \mathbf{J}_s, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\mathbf{J}_1 = \cdots = \mathbf{J}_s$ is satisfied for $N^d$ cases among $\mathbf{J}_1, \cdots, \mathbf{J}_s \in \mathbb{Z}_N^d$, we have

$$\begin{aligned} &\int_{\mathbb{T}^{sd}} h_1(\mathbf{x}_s) h_2(\mathbf{x}_s) \cdots h_s(\mathbf{x}_s) P_N(\mathbf{x}_1, \cdots, \mathbf{x}_s) \mathbf{dx}_1 \cdots \mathbf{dx}_s \\ &= (2\pi)^{(s-1)d} N^d \int_{\mathbb{T}^d} h_1(\mathbf{x}_s) h_2(\mathbf{x}_s) \cdots h_s(\mathbf{x}_s) \mathbf{dx}_s. \end{aligned}$$

□

Now we are ready to prove Theorem 4.1.

*Proof of Theorem 4.1.* We will investigate the limits of all the joint cumulants of

$$N^{d/2}\delta^{d-\alpha} \int_{\mathbb{T}^d} \mathbf{\Phi}(\boldsymbol{\omega}) \left( \mathcal{I}_{N,\delta}^{\tau}(\boldsymbol{\omega}) - \mathbb{E}\mathcal{I}_{N,\delta}^{\tau}(\boldsymbol{\omega}) \right) \mathbf{d}\boldsymbol{\omega}.$$

Since the first term is subtracted by its expectation, the first cumulant is obviously zero. Second cumulant between $k_1$th and $k_2$th components are given as

$$\begin{aligned} &\operatorname{cum}\left( N^{d/2}\delta^{d-\alpha} \int_{\mathbb{T}^d} \phi_{k_1}(\boldsymbol{\omega}) \mathcal{I}_{N,\delta}^{\tau}(\boldsymbol{\omega}) \mathbf{d}\boldsymbol{\omega}, N^{d/2}\delta^{d-\alpha} \int_{\mathbb{T}^d} \phi_{k_2}(\boldsymbol{\omega}) \mathcal{I}_{N,\delta}^{\tau}(\boldsymbol{\omega}) \mathbf{d}\boldsymbol{\omega} \right) \\ &= N^d \times (2\pi N)^{-2d} \int_{\mathbb{T}^{4d}} \{ \phi_{k_1}(\boldsymbol{\omega}_1) \phi_{k_2}(\boldsymbol{\omega}_2) + \phi_{k_1}(\boldsymbol{\omega}_1) \phi_{k_2}(-\boldsymbol{\omega}_2) \} \\ &\qquad \times \delta^{2(d-\alpha)} \bar{f}_{\delta}^{\tau}(\boldsymbol{\upsilon}_1) \bar{f}_{\delta}^{\tau}(\boldsymbol{\upsilon}_2) P_N(\boldsymbol{\omega}_1, \boldsymbol{\upsilon}_1, \boldsymbol{\omega}_2, \boldsymbol{\upsilon}_2) \mathbf{d}\boldsymbol{\omega}_1 \mathbf{d}\boldsymbol{\omega}_2 \mathbf{d}\boldsymbol{\upsilon}_1 \mathbf{d}\boldsymbol{\upsilon}_2 \end{aligned}$$

for all $k_1, k_2 \in \{1, \cdots, p\}$ by Lemma A.1. We can apply Lemma A.2 with $s = 4$ which tells us the difference between the above and

$$(2\pi)^d \int_{\mathbb{T}^d} \{\phi_{k_1}(\boldsymbol{\omega})\phi_{k_2}(\boldsymbol{\omega}) + \phi_{k_1}(\boldsymbol{\omega})\phi_{k_2}(-\boldsymbol{\omega})\} \, \delta^{2(d-\alpha)} \left(\bar{f}_\delta^\tau(\boldsymbol{\omega})\right)^2 \mathbf{d}\boldsymbol{\omega}$$

can be bounded by $\frac{(\log N)^{2d+1}}{N}$ times the combinations of $L^\infty$ norm of $\phi_{k_1}, \phi_{k_2}$, $\delta^{d-\alpha}\bar{f}_\delta^\tau$ and their derivatives. Since we assume the uniform convergence of $\delta^{d-\alpha}\bar{f}_\delta^\tau$ toward $g(\boldsymbol{\theta})$ with uniform convergence of its partial derivatives, we know that all the $L^\infty$ norms are bounded. Hence the difference goes to zero. Meanwhile from the uniform convergence, the above integral converges to

$$(2\pi)^d \int_{\mathbb{T}^d} \{\phi_{k_1}(\boldsymbol{\omega})\phi_{k_2}(\boldsymbol{\omega}) + \phi_{k_1}(\boldsymbol{\omega})\phi_{k_2}(-\boldsymbol{\omega})\} \, g^2(\boldsymbol{\omega}; \boldsymbol{\theta}) \mathbf{d}\boldsymbol{\omega}.$$

Now consider the $m$th order joint cumulant

$$\text{cum}\left(N^{d/2}\delta^{d-\alpha}\int_{\mathbb{T}^d} \phi_{k_1}(\boldsymbol{\omega})\mathcal{I}_{N,\delta}^\tau(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega}, \cdots N^{d/2}\delta^{d-\alpha}\int_{\mathbb{T}^d} \phi_{k_m}(\boldsymbol{\omega})\mathcal{I}_{N,\delta}^\tau(\boldsymbol{\omega})\mathbf{d}\boldsymbol{\omega}\right)$$

for $k_1, \cdots, k_m \in \{1, \cdots, p\}$ and $m \geq 3$. This is given as

$$N^{md/2} \times (2\pi N)^{-md} \sum_{(\boldsymbol{\iota}, \mathbf{e})} \int_{\mathbb{T}^{2md}} \phi_{k_1}(\boldsymbol{\omega}_1)\phi_{k_2}((-1)^{\bar{e}_2}\boldsymbol{\omega}_{\iota_2}) \cdots \phi_{k_m}((-1)^{\bar{e}_m}\boldsymbol{\omega}_{\iota_m})$$

$$\times \, \delta^{m(d-\alpha)} \bar{f}_\delta^\tau(\boldsymbol{\upsilon}_1) \cdots \bar{f}_\delta^\tau(\boldsymbol{\upsilon}_m) P_N(\boldsymbol{\omega}_1, \boldsymbol{\upsilon}_1, \cdots, \boldsymbol{\omega}_m, \boldsymbol{\upsilon}_m) \mathbf{d}\boldsymbol{\omega}_1 \cdots \mathbf{d}\boldsymbol{\omega}_m \mathbf{d}\boldsymbol{\upsilon}_1 \cdots \mathbf{d}\boldsymbol{\upsilon}_m$$

by Lemma A.1. Again by means of Lemma A.2, the difference between

$$\frac{\delta^{m(d-\alpha)}}{(2\pi)^{(2m-1)d} N^d} \int_{\mathbb{T}^{2md}} \phi_{k_1}(\boldsymbol{\omega}_1)\phi_{k_2}((-1)^{\bar{e}_2}\boldsymbol{\omega}_{\iota_2}) \cdots \phi_{k_m}((-1)^{\bar{e}_m}\boldsymbol{\omega}_{\iota_m}) \bar{f}_\delta^\tau(\boldsymbol{\upsilon}_1) \cdots \bar{f}_\delta^\tau(\boldsymbol{\upsilon}_m)$$

$$\times P_N(\boldsymbol{\omega}_1, \boldsymbol{\upsilon}_1, \cdots, \boldsymbol{\omega}_m, \boldsymbol{\upsilon}_m) \mathbf{d}\boldsymbol{\omega}_1 \cdots \mathbf{d}\boldsymbol{\omega}_m \mathbf{d}\boldsymbol{\upsilon}_1 \cdots \mathbf{d}\boldsymbol{\upsilon}_m$$

and

$$\int_{\mathbb{T}^d} \phi_{k_1}(\boldsymbol{\omega})\phi_{k_2}((-1)^{\bar{e}_2}\boldsymbol{\omega}) \cdots \phi_{k_m}((-1)^{\bar{e}_m}\boldsymbol{\omega}) \left(\delta^{d-\alpha}\bar{f}_\delta^\tau(\boldsymbol{\omega})\right)^m \mathbf{d}\boldsymbol{\omega}.$$

is bounded by a constant multiple of $\frac{(\log N)^{(2m-2)d+1}}{N}$ which goes to zero. Note that the above converges to

$$\int_{\mathbb{T}^d} \phi_{k_1}(\boldsymbol{\omega})\phi_{k_2}((-1)^{\bar{e}_2}\boldsymbol{\omega}) \cdots \phi_{k_m}((-1)^{\bar{e}_m}\boldsymbol{\omega}) \left(g(\boldsymbol{\omega};\boldsymbol{\theta})\right)^m \mathbf{d}\boldsymbol{\omega}.$$

Meanwhile

$$N^{md/2} \times (2\pi N)^{-md} \times (2\pi)^{(2m-1)d} N^d = (2\pi)^{(m-1)d} N^{-(m/2-1)d} \to 0$$

as $N \to \infty$ when $m \geq 3$. Hence we conclude that all the third and higher order cumulants goes to zero.

The claim for the expectation of fixed domain smoothed periodogram can also be proved in exactly the same way, by Lemma A.1 and Lemma A.2 with $s = 2$. □

# Appendix B

# Proofs for Chapter 5

Now we begin to prove Theorem 5.1. We first introduce the next two lemmas which gives a quantitative bound on the difference between the value of the function and its approximation by convolution with kernel functions.

**Lemma B.1.** *Suppose $h \in C^1(\mathbb{T}^d)$. $K_n$ represents the Fejér kernel of order $n$. Then*

$$\sup_{\boldsymbol{\omega}\in\mathbb{T}^d}\left|h(\boldsymbol{\omega})-\int_{\mathbb{T}^d}h(\boldsymbol{\upsilon})K_n(\boldsymbol{\omega}-\boldsymbol{\upsilon})\mathbf{d}\boldsymbol{\upsilon}\right|\lesssim\frac{\log n}{n}\|\partial h\|_{L^\infty}$$

*where*

$$\|\partial h\|_{L^\infty}=\sup\left\{\left|\frac{\partial h}{\partial\omega_j}(\boldsymbol{\omega})\right|:\boldsymbol{\omega}\in\mathbb{T}^d,j=1,\cdots,d\right\}.$$

*Proof.* We can write the left side of the statement in lemma as

$$\int_{\mathbb{T}^d}\left(h(\boldsymbol{\omega})-h(\boldsymbol{\omega}-\boldsymbol{\upsilon})\right)K_n(\boldsymbol{\upsilon})\mathbf{d}\boldsymbol{\upsilon}.$$

Since $h \in C^1(\mathbb{T}^d)$, we can bound

$$|h(\boldsymbol{\omega} - \boldsymbol{\upsilon}) - h(\boldsymbol{\omega})| \leq \|\partial h\|_{L^\infty} \sum_{j=1}^{d} |\upsilon_j|.$$

From

$$\begin{aligned}\int_{\mathbb{T}^d} |\upsilon_1| K_n(\boldsymbol{\upsilon}) \mathbf{d}\boldsymbol{\upsilon} &\lesssim \int_{\mathbb{T}} \frac{1}{n} \frac{|\upsilon_1|}{(n^{-1} + |\upsilon_1|)^2} d\upsilon_1 \times \prod_{j=2}^{d} \int_{\mathbb{T}} \frac{1}{2\pi n} \frac{\sin^2\left(\frac{n\upsilon_j}{2}\right)}{\sin^2\left(\frac{\upsilon_j}{2}\right)} d\upsilon_j \\ &\lesssim \frac{\log n}{n},\end{aligned}$$

we get the desired result. □

**Lemma B.2.** *Suppose $h \in C^1(\mathbb{T}^d)$. $K$ represents a kernel function satisfying $K \in C^1(\mathbb{T}^d), K \geq 0, K(\mathbf{x}) = K(-\mathbf{x}), \int K(\mathbf{x})\mathbf{dx} = 1$ and $K(\mathbf{x}) = 0$ whenever $\|\mathbf{x}\|_\infty \geq 1$. Denote the bandwidth as $b > 0$. Then*

$$\sup_{\boldsymbol{\omega} \in \mathbb{T}^d} \left| h(\boldsymbol{\omega}) - \int_{\mathbb{T}^d} h(\boldsymbol{\upsilon}) \frac{1}{b^d} K\left(\frac{\boldsymbol{\omega} - \boldsymbol{\upsilon}}{b}\right) \mathbf{d}\boldsymbol{\upsilon} \right| \leq b\|\partial h\|_{L^\infty}$$

*where $\|\partial h\|_{L^\infty}$ is defined as same as in Lemma B.1.*

*Proof.* The proof is essentially identical to that of Lemma B.1. All we need to find is the bound for $\int_{\mathbb{T}^d} |\upsilon_j| \frac{1}{b^d} K\left(\frac{\boldsymbol{\upsilon}}{b}\right) \mathbf{d}\boldsymbol{\upsilon}$ for each $j$. Since

$$\int_{\mathbb{T}^d} |\upsilon_j| \frac{1}{b^d} K\left(\frac{\boldsymbol{\upsilon}}{b}\right) \mathbf{d}\boldsymbol{\upsilon} = \int_{\mathbb{T}^d} b|\upsilon_j| K(\boldsymbol{\upsilon}) \mathbf{d}\boldsymbol{\upsilon} = \int_{[-1,1]^d} b|\upsilon_j| K(\boldsymbol{\upsilon}) \mathbf{d}\boldsymbol{\upsilon} \leq b \int_{[-1,1]^d} K(\boldsymbol{\upsilon}) \mathbf{d}\boldsymbol{\upsilon}$$

and the last term is equal to $b$, we get the desired result. □

Before directly dealing with $L_N(\boldsymbol{\theta})$, recall that smoothed periodograms need to be scaled by $\delta^{d-\alpha}$ to have a nondegenerate limit. This complicates theoretical investigation of $L_N(\boldsymbol{\theta})$ when the true value of $\alpha$ is unknown. Hence we choose a detour that assumes scaled tapered periodogram $\delta^{d-\alpha_0} \hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})$ to be known

when the true parameter is $\boldsymbol{\theta}_0 = (c_0, \alpha_0, \boldsymbol{\beta}_0)$. Be aware that this requires $\alpha_0$ but we still estimate $\alpha$ together with the other tail parameters in the following proposition. This seems unnatural but is needed to extend our method for the case where $\alpha$ is unknown.

**Proposition B.3.** *Assume the same as in Theorem 5.1 except Assumption 1. Define*

$$L_{0N}(\boldsymbol{\theta}) = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \left\{ \frac{\delta^{d-\alpha_0}\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})}{g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})} + \log g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}) \right\}$$

*and assume (A1), (A2-weak), (A3)-(A5) where (A2-weak) is the weaker version of (A2) which is*

*(A2-weak)* $\delta^{d-\alpha}\bar{f}^{\tau}_{\delta}(\boldsymbol{\omega})$ *uniformly converges to* $g(\boldsymbol{\omega};\boldsymbol{\theta}) \in C^1(\mathbb{T}^d)$ *with uniformly converging first derivatives for each* $\boldsymbol{\theta} \in \Theta$.

*Then there exists* $M_0 \in \mathbb{N}$ *such that for all* $M \geq M_0$, *the minimizer* $\hat{\boldsymbol{\theta}}$ *of* $L_{0N}(\boldsymbol{\theta})$ *converges to* $\boldsymbol{\theta}_0$ *in probability as* $N \to \infty$.

*Proof.* Theorem 4.1 can be applied due to the assumption (A2-weak) so that we have

$$\delta^{d-\alpha_0}\left(\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega}) - \bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})\right) = O_P(N^{-d/2})$$

for each $\boldsymbol{\omega} \in \mathbb{T}^d$ under the true measure. Moreover since $\|K_n\|_{L^1} = 1$ for all $n \in \mathbb{N}$ and $\left\|\frac{1}{h^d}K\left(\frac{\cdot}{h}\right)\right\|_{L^1} = 1$ for all $h > 0$, the assumption (A2-weak) and

Young's convolution inequality implies

$$\begin{aligned}
&\sup_{\boldsymbol{\omega}\in\mathbb{T}^d}\left|\delta^{d-\alpha_0}\bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})-g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}_0)\right| \\
&=\|\delta^{d-\alpha_0}\bar{f}^{\tau}_{N,\delta,M}-g_{N,M}(\boldsymbol{\theta}_0)\|_{L^\infty} \\
&=\|(\delta^{d-\alpha_0}\bar{f}^{\tau}_{\delta}-g(\boldsymbol{\theta}_0))*K_N*K_M\|_{L^\infty} \\
&=\|\delta^{d-\alpha_0}\bar{f}^{\tau}_{\delta}-g(\boldsymbol{\theta}_0)\|_{L^\infty}\|K_N\|_{L^1}\|K_M\|_{L^1}\to 0
\end{aligned}$$

as $N\to\infty$ and the same holds when $K_M$ is replaced by $\frac{1}{h^d}K\left(\frac{\cdot}{h}\right)$. From now on, we prove the result for the tapered periodogram. The proof is essentially identical for the smoothed periodogram with a compactly supported kernel.

Recall that $\boldsymbol{\theta}=(c,\alpha,\boldsymbol{\beta})$ and $g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta})=c\cdot g_{N,M}(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})$. We will now show that $g_{N,M}(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})\to g_M(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})$ for $\boldsymbol{\omega}\in\Omega_M$ and $(\alpha,\boldsymbol{\beta})\in\mathcal{A}\times\mathcal{B}$ uniformly as $N\to\infty$. Since $g_{N,M}(1,\alpha,\boldsymbol{\beta})=g_M(1,\alpha,\boldsymbol{\beta})*K_N$, we can consider to apply Lemma B.1. Beforehand we need to check whether $g_M(\cdot;1,\alpha,\boldsymbol{\beta})\in C^1(\mathbb{T}^d)$ is satisfied. From the expression

$$g_M(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})=\int_{\mathbb{T}^d}g(\boldsymbol{\omega}-\boldsymbol{\upsilon};1,\alpha,\boldsymbol{\beta})K_M(\boldsymbol{\upsilon})\mathbf{d}\boldsymbol{\upsilon},$$

we can see that $g_M(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})\in C^1(\mathbb{T}^d)$ is implied by $g(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})\in C^1(\mathbb{T}^d)$ and Leibniz rule which is allowed to be applied by means of the assumption (A3). Hence we can apply Lemma B.1 to attain

$$\begin{aligned}
|g_{N,M}(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})-g_M(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})|&\lesssim\frac{\log N}{N}\|\partial_{\boldsymbol{\omega}}g_M(1,\alpha,\boldsymbol{\beta})\|_{L^\infty} \\
&\leq\frac{\log N}{N}\|\partial_{\boldsymbol{\omega}}g(1,\alpha,\boldsymbol{\beta})\|_{L^\infty},
\end{aligned}$$

where the meaning of $\|\partial_{\boldsymbol{\omega}}h\|_{L^\infty}$ is the same as that in Lemma B.1. Subscripted notation clarifies the variable that derivatives are calculated. The last inequality

is from convolution inequality. Since the assumption (A3) again gives a uniform bound on $\|\partial_{\boldsymbol{\omega}} g(1,\alpha,\boldsymbol{\beta})\|_{L^\infty}$, we have

$$\sup\left\{|g_{N,M}(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta}) - g_M(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})| : \boldsymbol{\omega}\in\mathbb{T}^d, (\alpha,\boldsymbol{\beta})\in\mathcal{A}\times\mathcal{B}\right\} \lesssim \frac{\log N}{N} \tag{B.1}$$

which confirms our claim.

Observe that

$$L_{0N}(\boldsymbol{\theta}) = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M}\left\{\frac{1}{c}\frac{\delta^{d-\alpha_0}\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})}{g_{N,M}(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})} + \log c + \log g_{N,M}(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})\right\}.$$

Applying profiling approach, we can minimize $L_{0N}$ with respect to $c$ for each $(\alpha,\boldsymbol{\beta})$. The minimizer, denoting as $\hat{c}(\alpha,\boldsymbol{\beta})$, is given as

$$\hat{c}(\alpha,\boldsymbol{\beta}) = \frac{1}{|\Omega_M|}\sum_{\boldsymbol{\omega}\in\Omega_M}\frac{\delta^{d-\alpha_0}\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})}{g_{N,M}(\boldsymbol{\omega};1,\alpha,\boldsymbol{\beta})}$$

where $|\Omega_M|$ is the cardinality of $\Omega_M$. From (B.1) with the assumptions (A3) and (A4), we can take some closed interval $\mathcal{C}\in\mathbb{R}^+$ such that

$$\mathbb{P}\left(\hat{c}(\alpha,\boldsymbol{\beta})\in\mathcal{C}, \forall(\alpha,\boldsymbol{\beta})\in\mathcal{A}\times\mathcal{B}\right)\to 1$$

as $N\to\infty$. This implies that

$$\mathbb{P}\left(\arg\min_{\boldsymbol{\theta}\in\Theta} L_{0N}(\boldsymbol{\theta})\in\mathcal{C}\times\mathcal{A}\times\mathcal{B}\right)\to 1$$

as $N\to\infty$. Henceforth it is enough to consider minimization of $L_{0N}(\boldsymbol{\theta})$ over a compact set $\Theta' := \mathcal{C}\times\mathcal{A}\times\mathcal{B}$.

To show the consistency, observe that

$$\mathbb{P}\left(|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0|\geq\epsilon\right)\leq\mathbb{P}\left(\inf\left\{L_{0N}(\boldsymbol{\theta})-L_{0N}(\boldsymbol{\theta}_0) : |\boldsymbol{\theta}-\boldsymbol{\theta}_0|\geq\epsilon, \boldsymbol{\theta}\in\Theta'\right\}\leq 0\right)$$

for large enough $N$. Hence it is enough to show the right hand side of the above goes to zero. Define

$$L_0(\boldsymbol{\theta}) = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \left\{ \frac{g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g_M(\boldsymbol{\omega};\boldsymbol{\theta})} + \log g_M(\boldsymbol{\omega};\boldsymbol{\theta}) \right\}.$$

We have

$$L_{0N}(\boldsymbol{\theta}) - L_{0N}(\boldsymbol{\theta}_0) \geq L_0(\boldsymbol{\theta}) - L_0(\boldsymbol{\theta}_0) - 2\sup_{\boldsymbol{\theta}\in\Theta'} |L_{0N}(\boldsymbol{\theta}) - L_0(\boldsymbol{\theta})|$$

on $\boldsymbol{\theta} \in \Theta'$. Since (B.1) and the assumption (A4) implies

$$\sup_{\boldsymbol{\theta}\in\Theta'} |L_{0N}(\boldsymbol{\theta}) - L_0(\boldsymbol{\theta})| \xrightarrow[N\to\infty]{p} 0,$$

it is enough to show

$$\inf_{\boldsymbol{\theta}\in\Theta', |\boldsymbol{\theta}-\boldsymbol{\theta}_0|\geq\epsilon} \{L_0(\boldsymbol{\theta}) - L_0(\boldsymbol{\theta}_0)\} > 0 \tag{B.2}$$

for every $\epsilon > 0$. We have

$$\begin{aligned} L_0(\boldsymbol{\theta}) - L_0(\boldsymbol{\theta}_0) &= \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \left\{ \frac{g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g_M(\boldsymbol{\omega};\boldsymbol{\theta})} - 1 - \log \frac{g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g_M(\boldsymbol{\omega};\boldsymbol{\theta})} \right\} \\ &= \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \varphi\left( \frac{g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g_M(\boldsymbol{\omega};\boldsymbol{\theta})} \right) \end{aligned}$$

where $\varphi(x) = x - 1 - \log x$. Again using Lemma B.1 (Lemma B.2 for the compact support kernel version) with the assumptions (A3) and (A4), we can say that

$$\sup_{\boldsymbol{\omega}\in\mathbb{T}^d\setminus(-t,t)^d, \boldsymbol{\theta}\in\Theta} \left| \frac{g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g_M(\boldsymbol{\omega};\boldsymbol{\theta})} - \frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})} \right| \lesssim \frac{\log M}{M}.$$

The same conditions also implies that $\frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})}$ is bounded and away from zero uniformly in $\boldsymbol{\omega} \in \mathbb{T}^d \setminus (-t,t)^d$ and $\boldsymbol{\theta} \in \Theta$. Hence we can claim that

$$\left| \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \varphi\left( \frac{g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g_M(\boldsymbol{\omega};\boldsymbol{\theta})} \right) - \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \varphi\left( \frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})} \right) \right| \lesssim \frac{\log M}{M}.$$

Moreover since $\Omega_M$ is a subset of a regular grid whose neighboring distance is $\frac{2\pi}{M}$, we can consider the above term as Riemann sum approximation so that

$$\left| \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \varphi\left(\frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})}\right) - \int_{\mathbb{T}^d\setminus(-t,t)^d} \varphi\left(\frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})}\right) \mathbf{d}\boldsymbol{\omega} \right| \lesssim \frac{1}{M}.$$

Define $I(\boldsymbol{\theta})$ as

$$I(\boldsymbol{\theta}) = \int_{\mathbb{T}^d\setminus(-t,t)^d} \varphi\left(\frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})}\right) \mathbf{d}\boldsymbol{\omega}.$$

Since we can take $C > 0$ independently from any terms, especially $M$, satisfying $L_0(\boldsymbol{\theta}) - L_0(\boldsymbol{\theta}_0) \geq I(\boldsymbol{\theta}) - \frac{C\log M}{M}$, it is enough to show

$$\inf_{\boldsymbol{\theta}\in\Theta', |\boldsymbol{\theta}-\boldsymbol{\theta}_0|\geq\epsilon} I(\boldsymbol{\theta}) > 0 \tag{B.3}$$

to complete the proof. This is because (B.3) implies that (B.2) holds for all $M \geq M_0$ by taking $M_0$ to satisfy

$$\inf_{\boldsymbol{\theta}\in\Theta', |\boldsymbol{\theta}-\boldsymbol{\theta}_0|\geq\epsilon} \{I(\boldsymbol{\theta})\} - \frac{C\log M_0}{M_0} > 0.$$

Note that $\varphi \geq 0$ and $\varphi(x) = 0$ if and only if $x = 1$, which means that $I(\boldsymbol{\theta}) \geq 0$ where equality holds if and only if $g(\boldsymbol{\omega};\boldsymbol{\theta}_0) = g(\boldsymbol{\omega};\boldsymbol{\theta})$ almost everywhere on $\boldsymbol{\omega} \in \mathbb{T}^d \setminus (-t,t)^d$. Now, if $\boldsymbol{\theta} \mapsto I(\boldsymbol{\theta})$ is continuous, we can show (B.3). Indeed, from the continuity of $I(\boldsymbol{\theta})$ with compactness of $U = \Theta' \cap \{\boldsymbol{\theta} : |\boldsymbol{\theta}-\boldsymbol{\theta}_0| \geq \epsilon\}$, $L(\boldsymbol{\theta})$ is guaranteed to achieve minimum on $U$ so that there exists $\boldsymbol{\theta}' \in U$ satisfying $\inf_{\boldsymbol{\theta}\in U} I(\boldsymbol{\theta}) = I(\boldsymbol{\theta}')$. If $I(\boldsymbol{\theta}') = 0$, it means $g(\boldsymbol{\omega};\boldsymbol{\theta}') = g(\boldsymbol{\omega};\boldsymbol{\theta}_0)$ for almost every $\boldsymbol{\omega} \in \mathbb{T}^d \setminus (-t,t)^d$. Since $\boldsymbol{\theta}'$ cannot be $\boldsymbol{\theta}_0$, it contradicts to the identifiabililty

assumption (A1). The continuity of $I(\boldsymbol{\theta})$ can be shown by

$$\begin{aligned}|I(\boldsymbol{\theta}) - I(\tilde{\boldsymbol{\theta}})| &\leq \int_{\mathbb{T}^d \setminus (-t,t)^d} \left| \varphi\left(\frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})}\right) - \varphi\left(\frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\tilde{\boldsymbol{\theta}})}\right) \right| \mathbf{d}\boldsymbol{\omega} \\ &\leq \sup_{s \in [c_L, c_U]} |\varphi'(s)| \times \int_{\mathbb{T}^d \setminus (-t,t)^d} \left| \frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})} - \frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\tilde{\boldsymbol{\theta}})} \right| \mathbf{d}\boldsymbol{\omega}\end{aligned}$$

where $0 < c_L \leq \frac{g(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g(\boldsymbol{\omega};\boldsymbol{\theta})} \leq c_U < \infty$ for all $\boldsymbol{\omega} \in \mathbb{T}^d \setminus (-t,t)^d$ and $\boldsymbol{\theta} \in \Theta'$. This can be done by the assumptions (A3) and (A4). The same conditions again implies that the integrand in the right hand side is bounded. Now from the assumption (A5), we can apply dominant convergence theorem so that the right hand side goes to zero as $|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}| \to 0$. This shows the continuity of $I(\boldsymbol{\theta})$ which completes the proof. □

We have shown the consistency of $\hat{\boldsymbol{\theta}}$ in Proposition B.3 when $\alpha_0$ is known. If we assume (A2) which is stronger than (A2-weak), then we can also derive convergence order of the estimator. Moreover, asymptotic normality of the estimator can be derived as well when $\frac{d}{2} < \gamma$ where $\gamma$ is the constant that characterizes how fast $\delta^{d-\alpha} \bar{f}_\delta^\tau$ converges to $g(\boldsymbol{\theta})$ and is defined in the assumption (A2). These results are stated formally in the next proposition.

**Proposition B.4.** *Assume the same as in Proposition B.3 and additionally assume Assumption 1. Then there exists $M_0 \in \mathbb{N}$ such that for all $M \geq M_0$, the minimizer $\hat{\boldsymbol{\theta}}$ of $L_{0N}(\boldsymbol{\theta})$ satisfies*

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_P(N^{-\min\{\frac{d}{2},\gamma\}}).$$

*If $\frac{d}{2} < \gamma$ and the tapered periodogram is used, then*

$$N^{d/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow[N\to\infty]{d} N(\mathbf{0}, (2\pi)^d \mathcal{J}^{-1} \Sigma \mathcal{J}^{-1})$$

*where*

$$\mathcal{J} = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \left(\frac{\partial \log g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^T$$

*and*

$$\Sigma = \int_{\mathbb{T}^d} \left\{\boldsymbol{\Phi}_M(\mathbf{x})\boldsymbol{\Phi}_M^T(\mathbf{x}) + \boldsymbol{\Phi}_M(\mathbf{x})\boldsymbol{\Phi}_M^T(-\mathbf{x})\right\} g^2(\mathbf{x};\boldsymbol{\theta}_0)\mathbf{dx}$$

*with*

$$\boldsymbol{\Phi}_M(\mathbf{x}) = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} \frac{\partial \log g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{K_M(\boldsymbol{\omega}-\mathbf{x})}{g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}.$$

*When the smoothed periodogram with a compactly supported kernel is used, then the same result holds but replacing* $K_M$ *with* $\frac{1}{h^d}K\left(\frac{\cdot}{h}\right)$ *and* $g_M(\boldsymbol{\theta}_0)$ *with* $g_h(\boldsymbol{\theta}) = g(\boldsymbol{\theta}) * \frac{1}{h^d}K\left(\frac{\cdot}{h}\right)$*. In this case,* $\Sigma$ *gets simplified as*

$$\Sigma = (2h)^{2d} \sum_{\boldsymbol{\omega}\in\Omega_M} \left(\frac{\partial \log g_h(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log g_h(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^T \frac{1}{g_h^2(\boldsymbol{\omega};\boldsymbol{\theta}_0)} \times \int_{\mathbb{T}^d} g^2(\mathbf{x};\boldsymbol{\theta}_0)\left\{\frac{1}{h^d}K\left(\frac{\boldsymbol{\omega}-\mathbf{x}}{h}\right)\right\}^2 \mathbf{dx}.$$

*Proof.* We write the proof for the tapered periodogram. The whole arguments work identically for the smoothed periodogram with a compactly supported kernel.

Classical argument using Taylor expansion gives

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\left(\frac{\partial^2 L_{0N}(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\right)^{-1} \frac{\partial L_{0N}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}},$$

where $\boldsymbol{\theta}^*$ is a stochastic vector lies between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$ so that it converges to $\boldsymbol{\theta}_0$ in probability by means of Proposition B.3. This is actually an abuse of notation because $\boldsymbol{\theta}^*$ should be differently chosen as $\boldsymbol{\theta}^{k*}$ for each $k = 1, \cdots, l$ so

that $\frac{\partial L_{0N}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial L_{0N}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \sum_{k=1}^{l} \frac{\partial^2 L_{0N}(\boldsymbol{\theta}^{k*})}{\partial \boldsymbol{\theta} \partial \theta_k} (\hat{\theta}_k - \theta_{k0})$ should hold, however all of these $\boldsymbol{\theta}^{k*}$ converges to $\boldsymbol{\theta}_0$ in probability and this property is all we need. Thus we keep this abused notation below.

Direct calculation gives

$$\frac{\partial L_{0N}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega} \in \Omega_M} -\frac{\partial \log g_{N,M}(\boldsymbol{\omega}; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \left\{ \frac{\delta^{d-\alpha_0} \hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})}{g_{N,M}(\boldsymbol{\omega}; \boldsymbol{\theta}_0)} - 1 \right\}$$

and

$$\begin{aligned}\frac{\partial^2 L_{0N}(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} &= \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega} \in \Omega_M} -\frac{\partial^2 \log g_{N,M}(\boldsymbol{\omega}; \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \left\{ \frac{\delta^{d-\alpha_0} \hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})}{g_{N,M}(\boldsymbol{\omega}; \boldsymbol{\theta}^*)} - 1 \right\} \\ &+ \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega} \in \Omega_M} \left(\frac{\partial \log g_{N,M}(\boldsymbol{\omega}; \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}\right) \left(\frac{\partial \log g_{N,M}(\boldsymbol{\omega}; \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}\right)^T \frac{\delta^{d-\alpha_0} \hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})}{g_{N,M}(\boldsymbol{\omega}; \boldsymbol{\theta}^*)}.\end{aligned}$$

Lemma B.1 implies

$$\begin{aligned}|g_{N,M}(\boldsymbol{\omega}; 1, \alpha^*, \boldsymbol{\beta}^*) - g_M(\boldsymbol{\omega}; 1, \alpha^*, \boldsymbol{\beta}^*)| &\lesssim \frac{\log N}{N} \|\partial_{\boldsymbol{\omega}} g_M(1, \alpha^*, \boldsymbol{\beta}^*)\|_{L^\infty} \\ &\leq \frac{\log N}{N} \|\partial_{\boldsymbol{\omega}} g(1, \alpha^*, \boldsymbol{\beta}^*)\|_{L^\infty},\end{aligned}$$

and the assmptions (A3), (A5) and consistency of $\boldsymbol{\theta}^*$ implies

$$\begin{aligned}g_M(\boldsymbol{\omega}; 1, \alpha^*, \boldsymbol{\beta}^*) &= \int_{\mathbb{T}^d} g(\mathbf{x}; 1, \alpha^*, \boldsymbol{\beta}^*) K_M(\boldsymbol{\omega} - \mathbf{x}) \mathbf{dx} \\ &\xrightarrow[N \to \infty]{p} \int_{\mathbb{T}^d} g(\mathbf{x}; 1, \alpha_0, \boldsymbol{\beta}_0) K_M(\boldsymbol{\omega} - \mathbf{x}) \mathbf{dx} = g_M(\boldsymbol{\omega}; 1, \alpha_0, \boldsymbol{\beta}_0)\end{aligned}$$

from dominant convergence theorem. Similar arguments for first and second order derivatives for $g_{N,M}$ can be made because of the assumption (A6), from which it can be shown that

$$\frac{\partial^2 L_{0N}(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \xrightarrow[N \to \infty]{p} \left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega} \in \Omega_M} \left(\frac{\partial \log g_M(\boldsymbol{\omega}; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) \left(\frac{\partial \log g_M(\boldsymbol{\omega}; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^T.$$

Meanwhile Theorem 4.1 implies that

$$N^{d/2}\left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} -\frac{\partial \log g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\delta^{d-\alpha_0}\hat{\mathcal{I}}^{\tau}_{N,\delta,M}(\boldsymbol{\omega}) - \delta^{d-\alpha_0}\bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega})}{g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}_0)}$$

converges toward Gaussian distribution with mean zero and covariance matrix

$$\left(\frac{2\pi}{M}\right)^{2d} \int_{\mathbb{T}^d} \sum_{\boldsymbol{\omega}\in\Omega_M}\sum_{\boldsymbol{\upsilon}\in\Omega_M} \frac{\partial \log g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \log g_M(\boldsymbol{\upsilon};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} \frac{K_M(\boldsymbol{\omega}-\mathbf{x})K_M(\boldsymbol{\upsilon}-\mathbf{x})}{g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)g_M(\boldsymbol{\upsilon};\boldsymbol{\theta}_0)} g^2(\mathbf{x};\boldsymbol{\theta}_0)\mathbf{dx}$$

$$+\left(\frac{2\pi}{M}\right)^{2d} \int_{\mathbb{T}^d} \sum_{\boldsymbol{\omega}\in\Omega_M}\sum_{\boldsymbol{\upsilon}\in\Omega_M} \frac{\partial \log g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \log g_M(\boldsymbol{\upsilon};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} \frac{K_M(\boldsymbol{\omega}-\mathbf{x})K_M(\boldsymbol{\upsilon}+\mathbf{x})}{g_M(\boldsymbol{\omega};\boldsymbol{\theta}_0)g_M(\boldsymbol{\upsilon};\boldsymbol{\theta}_0)} g^2(\mathbf{x};\boldsymbol{\theta}_0)\mathbf{dx}.$$

The remaining term is

$$\left(\frac{2\pi}{M}\right)^d \sum_{\boldsymbol{\omega}\in\Omega_M} -\frac{\partial \log g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\delta^{d-\alpha_0}\bar{f}^{\tau}_{N,\delta,M}(\boldsymbol{\omega}) - g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}_0)}{g_{N,M}(\boldsymbol{\omega};\boldsymbol{\theta}_0)} = O(N^{-\gamma})$$

which follows from the assumption (A2). Combining two order equations, we get the desired result. □

Finally we are ready to prove Theorem 5.1, the version that $\alpha$ is unknown.

*Proof of Theorem 5.1.* Recall that $\boldsymbol{\theta} = (c, \alpha, \boldsymbol{\beta})$. We have

$$L_N(\boldsymbol{\theta}) = L_{0N}(c\delta^{\alpha-\alpha_0}, \alpha, \boldsymbol{\beta}) - \left(\frac{2\pi}{M}\right)^d |\Omega_M| \log \delta^{d-\alpha_0},$$

where $L_{0N}$ is that in Proposition B.3. Hence by denoting the local minimizer of $L_{0N}$ which is consistent to $\boldsymbol{\theta}_0$ as $\tilde{\boldsymbol{\theta}} = (\tilde{c}, \tilde{\alpha}, \tilde{\boldsymbol{\beta}})$, we have $(\hat{c}\delta^{\hat{\alpha}-\alpha_0}, \hat{\alpha}, \hat{\boldsymbol{\beta}}) = (\tilde{c}, \tilde{\alpha}, \tilde{\boldsymbol{\beta}})$ where $\hat{\boldsymbol{\theta}} = (\hat{c}, \hat{\alpha}, \hat{\boldsymbol{\beta}})$ is a local minimizer of $L_N$. The result in Proposition B.4 then implies the desired result for $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$. To derive the consistency result and convergence order for $\hat{c}$, we use $N^{\min\{\frac{d}{2},\gamma\}}(\hat{c}\delta^{\hat{\alpha}-\alpha_0} - c_0) = O_P(1)$ which comes from the same proposition. Hence we have

$$\log \hat{c} + (\hat{\alpha} - \alpha_0)\log\delta - \log c_0 = O_P(N^{-\min\{\frac{d}{2},\gamma\}}),$$

which gives us the desired results. □

# Appendix C

# Tables for Chapter 8

| $\tau = 1$ | $M = 25$ | $M = 50$ | $M = 100$ | $M = 200$ | Naive Average |
|---|---|---|---|---|---|
| sc98drft.002 | $\alpha = 2.35$ | $\alpha = 2.24$ | $\alpha = 2.19$ | $\alpha = 2.15$ | $\alpha = 2.23$ |
| | $D = 1.33$ | $D = 1.38$ | $D = 1.40$ | $D = 1.42$ | $D = 1.38$ |
| sc98drft.003 | $\alpha = 2.50$ | $\alpha = 2.42$ | $\alpha = 2.36$ | $\alpha = 2.32$ | $\alpha = 2.40$ |
| | $D = 1.25$ | $D = 1.29$ | $D = 1.32$ | $D = 1.34$ | $D = 1.30$ |
| sc98drft.005 | $\alpha = 2.44$ | $\alpha = 2.40$ | $\alpha = 2.37$ | $\alpha = 2.34$ | $\alpha = 2.39$ |
| | $D = 1.28$ | $D = 1.30$ | $D = 1.32$ | $D = 1.33$ | $D = 1.31$ |
| sc98drft.006 | $\alpha = 2.32$ | $\alpha = 2.29$ | $\alpha = 2.27$ | $\alpha = 2.24$ | $\alpha = 2.28$ |
| | $D = 1.34$ | $D = 1.35$ | $D = 1.37$ | $D = 1.38$ | $D = 1.36$ |
| sc98drft.007 | $\alpha = 2.35$ | $\alpha = 2.28$ | $\alpha = 2.20$ | $\alpha = 2.15$ | $\alpha = 2.24$ |
| | $D = 1.32$ | $D = 1.36$ | $D = 1.40$ | $D = 1.43$ | $D = 1.38$ |
| sc98drft.008 | $\alpha = 2.74$ | $\alpha = 2.67$ | $\alpha = 2.59$ | $\alpha = 2.55$ | $\alpha = 2.64$ |
| | $D = 1.13$ | $D = 1.17$ | $D = 1.20$ | $D = 1.23$ | $D = 1.18$ |

Table C.1 Estimated $\alpha$ and fractal dimension of 6 sea ice profiles based on tail estimation method under $\tau = 1$. The right most column is achieved by simple average of four estimates.

Moving window estimates for $\log c$

| lat \ lon | -120 | -115 | -110 | -105 | -100 | -95 |
|---|---|---|---|---|---|---|
| 60 | | | | | | |
| | 7.907 | 6.333 | 5.442 | 4.951 | 4.735 | 3.955 | 3.182 | 2.918 | 2.823 |
| 55 | 5.519 | 6.802 | 6.132 | 4.228 | 3.169 | 2.485 | 3.776 | 4.220 | 3.353 |
| | 3.090 | 2.327 | 5.457 | 4.402 | 2.704 | 0.390 | 3.741 | 4.691 | 3.858 |
| 50 | 4.058 | 3.480 | 5.656 | 4.365 | 4.011 | 3.005 | 1.846 | 1.864 | 0.659 |
| | 5.878 | 6.616 | 6.784 | 5.870 | 4.895 | 4.479 | 3.101 | 0.000 | -1.668 |
| 45 | 5.296 | 5.860 | 7.108 | 7.030 | 6.292 | 4.839 | 3.378 | 1.660 | -1.270 |
| | 6.319 | 5.590 | 6.840 | 7.196 | 7.005 | 5.755 | 0.911 | -0.047 | -1.947 |
| 40 | 8.455 | 7.334 | 7.208 | 6.783 | 6.530 | 6.584 | 0.849 | -3.829 | -3.627 |
| | 9.086 | 8.451 | 7.842 | 7.387 | 6.886 | 6.929 | 1.842 | -3.263 | -3.449 |
| 35 | | | | | | |

Table C.2 Moving window estimation result of $\log c$ for monthly maximum temperature data.

Moving window estimates for $\alpha$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 60 | | | | | | | | | |
| | 5.416 | 5.132 | 4.965 | 4.829 | 4.746 | 4.606 | 4.480 | 4.459 | 4.443 |
| 55 | 4.686 | 5.161 | 5.088 | 4.724 | 4.520 | 4.354 | 4.598 | 4.701 | 4.552 |
| | 3.907 | 3.880 | 4.953 | 4.768 | 4.432 | 3.924 | 4.610 | 4.795 | 4.656 |
| 50 | 4.136 | 4.112 | 4.921 | 4.733 | 4.675 | 4.471 | 4.252 | 4.253 | 4.000 |
| | 4.799 | 4.979 | 5.084 | 4.953 | 4.791 | 4.724 | 4.459 | 3.862 | 3.522 |
| 45 | 4.768 | 4.840 | 5.127 | 5.079 | 4.977 | 4.771 | 4.490 | 4.192 | 3.593 |
| | 5.083 | 4.871 | 5.102 | 5.132 | 5.128 | 4.961 | 4.011 | 3.841 | 3.447 |
| 40 | 5.451 | 5.306 | 5.198 | 5.042 | 4.851 | 4.929 | 4.037 | 3.072 | 3.094 |
| | 5.321 | 5.300 | 5.182 | 5.139 | 4.907 | 4.961 | 4.183 | 3.163 | 3.110 |
| 35 | | | | | | | | | |
| lat / lon | -120 | | -115 | | -110 | | -105 | | -100 | | -95 |

Table C.3 Moving window estimation result of $\alpha$ for monthly maximum temperature data.

Moving window estimates for $A_{11}$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 60 | | | | | | | | | |
| | 0.812 | 0.852 | 0.850 | 0.825 | 0.816 | 0.840 | 0.889 | 0.876 | 0.831 |
| 55 | 0.834 | 0.844 | 0.847 | 0.851 | 0.873 | 0.877 | 0.894 | 0.893 | 0.860 |
| | 0.915 | 0.901 | 0.858 | 0.844 | 0.854 | 0.847 | 0.886 | 0.903 | 0.906 |
| 50 | 0.969 | 0.981 | 0.925 | 0.840 | 0.853 | 0.865 | 0.900 | 0.927 | 0.937 |
| | 0.941 | 0.946 | 0.937 | 0.908 | 0.886 | 0.851 | 0.856 | 0.926 | 0.929 |
| 45 | 0.910 | 0.927 | 0.938 | 0.905 | 0.890 | 0.897 | 0.871 | 0.910 | 0.921 |
| | 0.964 | 0.959 | 0.971 | 0.913 | 0.903 | 0.949 | 0.900 | 0.891 | 0.880 |
| 40 | 0.965 | 1.033 | 1.008 | 0.929 | 0.922 | 0.937 | 0.903 | 0.788 | 0.823 |
| | 0.970 | 1.022 | 0.988 | 0.928 | 0.951 | 0.962 | 0.890 | 0.820 | 0.847 |
| 35 | | | | | | | | | |
| lat / lon | -120 | | -115 | | -110 | | -105 | | -100 | | -95 |

Table C.4 Moving window estimation result of $A_{11}$ for monthly maximum temperature data.

Moving window estimates for $A_{12}$

| lat \ lon | -120 | -115 | -110 | -105 | -100 | -95 |
|---|---|---|---|---|---|---|
| 60 | | | | | | |
| | 0.016 | -0.015 | -0.026 | -0.006 | -0.018 | -0.046 | -0.009 | -0.030 | -0.041 |
| 55 | 0.035 | 0.017 | 0.038 | 0.068 | 0.032 | -0.022 | -0.006 | -0.018 | -0.037 |
| | 0.107 | 0.264 | 0.145 | 0.144 | 0.130 | 0.070 | 0.031 | 0.027 | 0.012 |
| 50 | 0.052 | 0.188 | 0.156 | 0.118 | 0.028 | 0.073 | 0.148 | 0.123 | 0.099 |
| | 0.029 | 0.069 | 0.064 | -0.010 | -0.032 | 0.049 | 0.190 | 0.194 | 0.111 |
| 45 | -0.001 | 0.037 | 0.082 | 0.023 | 0.007 | 0.062 | 0.158 | 0.179 | 0.139 |
| | -0.046 | 0.031 | 0.112 | 0.049 | 0.020 | 0.033 | 0.094 | 0.148 | 0.166 |
| 40 | -0.040 | -0.073 | -0.012 | -0.008 | 0.062 | 0.077 | 0.027 | 0.242 | 0.202 |
| | 0.055 | -0.005 | -0.064 | -0.069 | 0.024 | 0.036 | 0.046 | 0.179 | 0.232 |
| 35 | | | | | | |

Table C.5 Moving window estimation result of $A_{12}$ for monthly maximum temperature data.

# Bibliography


[1] John T Abatzoglou, Solomon Z Dobrowski, Sean A Parks, and Katherine C Hegewisch. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific data*, 5(1): 1–12, 2018.

[2] Ethan Anderes. On the consistent separation of scale and variance for gaussian random fields. *The Annals of Statistics*, 38(2):870–893, 2010.

[3] Moreno Bevilacqua, Tarik Faouzi, Reinhard Furrer, and Emilio Porcu. Estimation and prediction using generalized wendland covariance functions under fixed domain asymptotics. *The Annals of Statistics*, 47(2):828–856, 2019.

[4] David R Brillinger. *Time series: data analysis and theory*. SIAM, 2001.

[5] Juan Du, Hao Zhang, and VS2549562 Mandrekar. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *the Annals of Statistics*, 37(6A):3330–3361, 2009.

[6] Montserrat Fuentes. Approximate likelihood for large irregularly spaced

spatial data. *Journal of the American Statistical Association*, 102(477): 321–331, 2007.

[7] Tilmann Gneiting, Hana Ševčíková, and Donald B Percival. Estimators of fractal dimension: Assessing the roughness of time series and spatial data. *Statistical Science*, pages 247–277, 2012.

[8] Joseph Guinness. Spectral density estimation for random fields via periodic embeddings. *Biometrika*, 106(2):267–286, 2019.

[9] Xavier Guyon. Parameter estimation for a stationary process on ad-dimensional lattice. *Biometrika*, 69(1):95–105, 1982.

[10] CC Heyde and R Gay. Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. *Stochastic Processes and their Applications*, 45(1):169–182, 1993.

[11] Cari G Kaufman, Mark J Schervish, and Douglas W Nychka. Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555, 2008.

[12] Chae Young Lim and Michael Stein. Properties of spatial cross-periodograms using fixed-domain asymptotics. *Journal of multivariate analysis*, 99(9):1962–1984, 2008.

[13] Wei-Liem Loh, Saifei Sun, and Jun Wen. On fixed-domain asymptotics, parameter estimation and isotropic gaussian random fields with matérn covariance functions. *The Annals of Statistics*, 49(6):3127–3152, 2021.

[14] Yasumasa Matsuda and Yoshihiro Yajima. Fourier analysis of irregularly spaced data on rd. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1):191–217, 2009.

[15] Camil Muscalu and Wilhelm Schlag. *Classical and Multilinear Harmonic Analysis: Volume 1*, volume 137. Cambridge University Press, 2013.

[16] Suhasini Subba Rao. Statistical inference for spatial statistics defined in the fourier domain. *The Annals of Statistics*, 46(2):469–499, 2018.

[17] Suhasini Subba Rao and Junho Yang. Reconciling the gaussian and whittle likelihood with an application to estimation in the frequency domain. *The Annals of Statistics*, 49(5):2774–2802, 2021.

[18] Peter M Robinson. Gaussian semiparametric estimation of long range dependence. *The Annals of statistics*, pages 1630–1661, 1995.

[19] Michael L Stein. Fixed-domain asymptotics for spatial periodograms. *Journal of the American Statistical Association*, 90(432):1277–1288, 1995.

[20] Michael L Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 1999.

[21] Michael L Stein. Equivalence of gaussian measures for some nonstationary random fields. *Journal of Statistical Planning and Inference*, 123(1):1–11, 2004.

[22] Adam M Sykulski, Sofia C Olhede, Arthur P Guillaumin, Jonathan M

Lilly, and Jeffrey J Early. The debiased whittle likelihood. *Biometrika*, 106(2):251–266, 2019.

[23] Daqing Wang and Wei-Liem Loh. On fixed-domain asymptotics and covariance tapering in gaussian random field models. *Electronic Journal of Statistics*, 5:238–269, 2011.

[24] Peter Whittle. The analysis of multiple stationary time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 15(1):125–139, 1953.

[25] Wei-Ying Wu and Chae Young Lim. Estimation of smoothness of a stationary gaussian random field. *Statistica Sinica*, pages 1729–1745, 2016.

[26] Wei-Ying Wu, Chae Young Lim, and Yimin Xiao. Tail estimation of the spectral density for a stationary gaussian random field. *Journal of Multivariate Analysis*, 116:74–91, 2013.

[27] Zhiliang Ying. Asymptotic properties of a maximum likelihood estimator with data from a gaussian process. *Journal of Multivariate analysis*, 36(2): 280–296, 1991.

[28] Zhiliang Ying. Maximum likelihood estimation of parameters under a spatial sampling scheme. *The Annals of Statistics*, pages 1567–1590, 1993.

[29] Hao Zhang. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261, 2004.

# 초록

정상 가우스 무작위장에 대한 고정 도메인 점근론에서는 스펙트럼 밀도함수의 꼬리행동의 역할을 강조하는 결과가 많다. 무작위장의 평활도, 잘못 지정된 스펙트럴 밀도함수를 통한 점근적 최적 크리깅, 그리고 가우시안 척도 등이 그 예시이다. 본 논문은 스펙트럴 밀도함수에 대한 꼬리 행동 모수 모형을 가정하였을 때, 고정 도메인 점근법 하에서 꼬리 행동 모수에 대한 일치추정량을 제안하는 것을 목표로 한다. 구체적으로, $d$차원 유클리드 공간의 유계 부분 집합 위의 정규 격자에서 수집된 데이터에 초점을 맞추었으며, 스펙트럴 영역에서의 분석법을 개발하기 위해 피리오도그램을 기반으로 Whittle 가능도 함수와 유사한 목적함수를 구성하였다. 이 과정에서 평활화된 피리오도그램이 중요하게 등장하므로 고정 도메인 점근법에서 평활화된 피리오도그램의 일치성과 점근정규성을 먼저 증명하였다. 다음으로, 테이퍼드 피리오도그램과 옹골 지지 커널로 평활화된 피리오도그램이라는 두 종류의 평활화된 피리오도그램을 기반으로, Whittle 가능도 함수와 유사한 목적함수를 구성하여 이를 최소로 만드는 해를 스펙트럴 꼬리 행동 모수의 추정량이 되도록 하였다. 추정량의 일치성과 수렴 속도, 그리고 일부 경우에서의 점근정규성을 증명하였다. 이 결과를 통해 추정량에 대한 통계적 추론이 가능해진다. 이론적 결과를 뒷받침하기 위한 시뮬레이션 실험을 진행하였으며, 현실 상황에서의 응용 예시로써 해빙 프로파일 데이터와 월간 최대 온도 데이터를 분석하여 제시하였다.

**주요어**: 고정 도메인 점근법, 가우시안 무작위장, 피리오도그램, 스펙트럴 밀도함수

**학번**: 2016-20269