Electronic Journal of Statistics
Vol. 11 (2017) 4297–4322
ISSN: 1935-7524
DOI: 10.1214/17-EJS1326

# An asymptotic theory for spectral analysis of random fields

**Soudeep Deb**

*Department of Statistics*
*University of Chicago*
*5747 S Ellis Ave*
*Chicago, IL 60637*
*e-mail:* sdeb@uchicago.edu

**Mohsen Pourahmadi**

*Department of Statistics*
*Texas A&M University*
*3143 TAMU*
*College Station, TX 77843-3143*
*e-mail:* pourahm@stat.tamu.edu

**and**

**Wei Biao Wu**

*Department of Statistics*
*University of Chicago*
*5747 S Ellis Ave*
*Jones 317*
*Chicago, IL 60637*
*e-mail:* wbwu@galton.uchicago.edu

**Abstract:** For a general class of stationary random fields we study asymptotic properties of the discrete Fourier transform (DFT), periodogram, parametric and nonparametric spectral density estimators under an easily verifiable short-range dependence condition expressed in terms of functional dependence measures. We allow irregularly spaced data which is indexed by a subset $\Gamma$ of $\mathbb{Z}^d$. Our asymptotic theory requires minimal restriction on the index set $\Gamma$. Asymptotic normality is derived for kernel spectral density estimators and the Whittle estimator of a parameterized spectral density function. We also develop asymptotic results for a covariance matrix estimate.

**Keywords and phrases:** Time series, random fields, irregular spaced data, spectral density.

Received March 2017.

## 1. Introduction

Analysis of irregularly spaced data has been attracting considerable attention from researchers in various fields, ranging from environmental science to economics. The origin of irregular data is in fact the limit theorem for random fields with continuous parameter where the sets of integration in the limit theorems approach to infinity in Van Hove sense, see for example, Ivanov and Leonenko (2012). In a broad sense, there are two different approaches to deal with the irregularly spaced spatial data. The classical and more popular Kriging or interpolation approach (Cressie (1988)) is parametric in nature. A nonparametric or a frequency domain approach was considered by Fuentes (2007) which revolves around the assumption that the sampled locations are fixed and not random. Vidal-Sanz (2009) also considered nonparametric estimation of spectral densities for second-order stationary random fields on a $d$-dimensional lattice. In that paper, the author proposed modified estimator classes with improved bias convergence rate. In a much recent work, Bandyopadhyay et al. (2015) formulated a spatial frequency domain empirical likelihood method for irregularly spaced data. Other works in the frequency domain can be found in Hall and Patil (1994), Bandyopadhyay and Lahiri (2009) and the references therein.

Spectral domain methods to approximate the Gaussian likelihood for irregularly spaced datasets were proposed by Matsuda and Yajima (2009) where the sampled locations are assumed random with a particular distribution having a continuous density function. Their non-parametric and parametric estimators of the spectral density function of the underlying random fields are similar to those in classical time series analysis. The parametric spectral density was estimated by minimizing the Whittle likelihood while the non-parametric spectral density estimator was a spectral window estimator, and they studied the asymptotic properties of those estimators.

In spatial data analysis one usually deals with irregularly spaced data. To set the notation, let $(\Gamma_n)_{n\geq 1}$ be a sequence of finite subsets of $\mathbb{Z}^d$ representing the sampling locations or design points. Our goal here is to work with an asymptotic regime which imposes minimal restrictions on the sampling set and its boundary:

**Assumption 1. (Asymptotic regime)** *Let* $\Gamma_n = \{L_{n,1}, \ldots, L_{n,n}\} \subset \mathbb{Z}^d$ *be the set of sampling locations such that the choice of* $\Gamma_n$ *satisfies the property* $|\Gamma_n| \to \infty$.

For simplicity, from now on, we would write $L_k = L_{n,k}$. It is instructive to compare the above regime with the Matsuda and Yajima (2009) setup where each sampling location $\mathbf{t}_i$ is obtained from a randomly generated $d$-dimensional vector $\mathbf{u}_i = (u_{i,1}, \ldots, u_{i,d})$ by $t_{ij} = A_j u_{ij}$ for $j = 1, 2, \ldots, d$. They further assumed that the coefficient $A_j$'s and the sample size $(n_k)$, if expressed as a function of $k$, satisfy the condition $|S_k| / n_k \to 0$ as $k$ goes to infinity, where $|S_k|$ is the area of the rectangle $[0, A_1] \times \ldots \times [0, A_d]$.

Other examples with special constraints on the sample set can be found in Jones (1962), Neave (1970a), Neave (1970b), Parzen (1963), Clinger and Van Ness (1976). A related study with irregularly spaced observations for an increasing spatio-temporal domain can be found in Li et al. (2008). Similar to Matsuda and Yajima (2009), they also viewed the spatial locations at which the data is observed as random in number and location; generated from a homogeneous 2-dimensional Poisson process. As an interesting feature, our theory requires a minimal condition on the set $\Gamma_n$, which is an attractive property in spatial applications in which the underlying observation domains can be quite irregular.

For a given $\Gamma_n$, the discrete Fourier transform (DFT) is defined by

$$S_n(\theta) = \frac{1}{\sqrt{n}} \sum_{j \in \Gamma_n} X_j \exp(-\imath j'\theta) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_{L_k} \exp(-\imath L_k'\theta), \tag{1.1}$$

where $\imath = \sqrt{-1}$ and $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$. The periodogram of the data is defined by

$$I_n(\theta) = |S_n(\theta)|^2 = \frac{1}{n}\left[\left\{\sum_{j \in \Gamma_n} X_j \cos(j'\theta)\right\}^2 + \left\{\sum_{j \in \Gamma_n} X_j \sin(j'\theta)\right\}^2\right]. \tag{1.2}$$

We will study aspects of both parametric and nonparametric estimators of the spectral density functions of stationary random fields. We begin by finding asymptotic results for the discrete Fourier transform (DFT) for irregularly spaced random fields, and then study the asymptotic properties of the spectral density estimates. As pointed out earlier, an important feature of our approach is that we do not impose any restriction on the index set $\Gamma_n$, other than the natural requirement $|\Gamma_n| \to \infty$.

The paper is structured as follows: Section 2 presents the setup, assumptions and some preliminary results regarding short-memory stationary random fields. The discrete Fourier transform of the data and its asymptotic properties are presented in Section 3 while the Whittle likelihood and parametric spectral density estimator are discussed in Section 4. In Section 5, we will present the nonparametric spectral density estimator and the covariance function and its different aspects (e.g. consistency, asymptotic normality). We will also discuss the estimation of covariances matrices for an irregular set-up in Section 6. All proofs are provided in Appendix.

## 2. Short-range dependent random fields

We shall consider a very general class of stationary random fields which are functions of independent and identically distributed (iid) random variables. In related works, Whittle (1954) considered two-dimensional linear auto-regression fields and Besag (1974) discussed stationary auto-normal processes and proposed estimation methods and goodness-of-fit tests applicable to spatial Markov schemes defined over a rectangular lattice. Other noteworthy examples may be found in Guyon (1982) and Kashyap (1984). Our setup is general enough to include the most common linear and nonlinear processes.

**Assumption 2.** *Let $\varepsilon_j, j \in \mathbb{Z}^d$, be iid random variables. Define*

$$X_i = g(\varepsilon_{i-s}; s \in \mathbb{Z}^d), \qquad i \in \mathbb{Z}^d, \tag{2.1}$$

*where $g$ is a measurable function such that $X_i$ is well-defined.*

Throughout the paper, we work with short-range dependent stationary processes. We use the idea of coupling (Wu (2005)) to define dependence measures. Let $\varepsilon_i', \varepsilon_j, i, j \in \mathbb{Z}^d$ be iid.

**Definition 2.1.** *(Functional dependence measure) Let $X_i \in Ł_p$, $p \geq 1$. Define*

$$\delta_{i,p} = \|X_i - X_i^*\|_p, \text{ where } X_i^* = g(\varepsilon_{i-s}^*; s \in \mathbb{Z}^d), \tag{2.2}$$

*and $\varepsilon_j^* = \varepsilon_j$ if $j \neq 0$ and $\varepsilon_0^* = \varepsilon_0'$. Also let $p' = \min\{2, p\}$ and define*

$$\Theta_{m,p} = \sum_{|j|>m} \delta_{j,p} \qquad \text{and} \qquad \Psi_{m,p} = \left( \sum_{|j|>m} \delta_{j,p}^{p'} \right)^{1/p'}.$$

**Definition 2.2.** *(Stability) The random field $(X_i)$ defined in (2.1) is said to be $p$-stable if*

$$\Delta_p := \sum_{i \in \mathbb{Z}^d} \delta_{i,p} < \infty.$$

Two concrete examples of such processes are given next.

**Example 1.** (Linear process) Let $X_i = \sum_{s \in \mathbb{Z}^d} a_s \varepsilon_{i-s}$, where $(\varepsilon_j)_{j \in \mathbb{Z}^d}$ are iid random variables with mean 0 and $\varepsilon_0 \in Ł_p$, $p \geq 2$, and $a_s$ are real coefficients such that $\sum_{s \in \mathbb{Z}^d} |a_s| < \infty$. Then $\delta_{i,p} = |a_i| \|\varepsilon_0 - \varepsilon_0^*\|_p = O(|a_i|)$. For the nonlinearly transformed process $Y_i = K(X_i)$, where $K(\cdot)$ is Lipschitz continuous, its functional dependence measure $\delta_{i,p}(Y)$ is also of order $O(|a_i|)$. □

**Example 2.** (Spatial Autoregressive Scheme) Let $\mathcal{N} \subset \mathbb{Z}^d$ be a finite set and $0 \notin \mathcal{N}$. Consider the spatial process in the form of nonlinear autoregressive scheme

$$X_i = G((X_{i-j})_{j \in \mathcal{N}}; \varepsilon_i),$$

where the function $G$ is such that there exists nonnegative numbers $\ell_j, j \in \mathcal{N}$, with $\sum_{j \in \mathcal{N}} \ell_j < 1$ and the following holds: for all $(x_{-j})_{j \in \mathcal{N}}$ and $(x'_{-j})_{j \in \mathcal{N}}$,

$$|G((x_{-j})_{j \in \mathcal{N}}; \varepsilon_i) - G((x'_{-j})_{j \in \mathcal{N}}; \varepsilon_i)| \leq \sum_{j \in \mathcal{N}} \ell_j |x_{-j} - x'_{-j}|. \tag{2.3}$$

Also assume that there exists $(x_{-j})_{j \in \mathcal{N}}$ such that $G((x_{-j})_{j \in \mathcal{N}}; \varepsilon_0) \in \mathcal{L}^p$. Then following the argument in Shao and Wu (2004), we have $\delta_{i,p} = O(\rho^{|i|})$ for some $0 < \rho < 1$. □

## 3. Asymptotic theory of the DFT

In this section, under some regularity conditions on the data-generating process we study the asymptotic distribution of DFT. These are the key ingredients for developing the spectral analysis of stationary processes. Peligrad and Wu (2010) proved a central limit theorem for the Fourier transform of a stationary process in the regular case. Other works related to bias and variance of periodogram estimates for a regular set-up can be found in Pukkila (1979) and Lin and Liu (2009).

### *3.1. Asymptotic normality of the DFT*

We establish the asymptotic normality of the DFT defined in (1.1) using the Cramer-Wold device. To this end, instead of the DFT we consider the more general expression

$$W_n = \sum_{j \in \Gamma_n} c_j X_j \qquad \text{where } |c_j| \le 1 \text{ for all } j \in \Gamma_n, \tag{3.1}$$

and wish to find the asymptotic joint distribution of $(Y_n(\theta), Z_n(\theta))/\sqrt{n}$ where

$$Y_n(\theta) = \sum_{j \in \Gamma_n} X_j \cos(j'\theta), \qquad Z_n(\theta) = -\sum_{j \in \Gamma_n} X_j \sin(j'\theta),$$

are the cosine and sine transforms of the data. As mentioned earlier, we are interested in linear combinations $aY_n(\theta) + bZ_n(\theta)$ (without loss of generality, we can assume that $a^2 + b^2 = 1$) which is of the form (3.1). For $k \ge 1$, let us use $N_k(0, \Sigma)$ to denote the $k$-variate normal distribution with 0 mean vector and variance-covariance matrix $\Sigma$.

**Theorem 3.1.** *(Central limit theorem for DFT). Suppose $(X_i)_{i \in \mathbb{Z}^d}$ is a stationary centered random field defined by (2.1) satisfying*

$$\Delta_2 := \sum_{i \in \mathbb{Z}^d} \delta_{i,2} < \infty. \tag{3.2}$$

*Also assume that $v_n^2 = \mathbb{E}(W_n^2) \to \infty$. Then, the following central limit theorem (CLT) holds:*

$$L\left[W_n/\sqrt{n}, N\big(0, v_n^2/n\big)\right] \to 0 \text{ as } n \to \infty, \tag{3.3}$$

*where $L(.,.)$ is the Levy distance between distributions. Consequently, we have*

$$L\left[\big(Y_n(\theta), Z_n(\theta)\big)/\sqrt{n}, N_2\big(0, \Sigma_n(\theta)/n\big)\right] \to 0, \tag{3.4}$$

*where $\Sigma_n(\theta) = \mathrm{cov}((Y_n(\theta), Z_n(\theta))^T$.*

The above theorem gives the joint asymptotic distribution of the discrete Fourier transform. For the regular set-up, we can further show that the two coordinates $(Y_n(\theta), Z_n(\theta))$ are asymptotically independent and the same will happen at two different frequencies.

**Proposition 1.** *Consider the regular set-up $\Gamma_n = \prod_{l=1}^d \{1, 2, \ldots, n_l\}$, with $n_l \to \infty$ for all $l \le d$. Let $\theta, \phi \in [-\pi, \pi)^d$ with $\theta, \phi \ne 0$, $\theta \ne \phi$ and $\theta + \phi \ne 0$. Then $(Y_n(\theta), Z_n(\theta), Y_n(\phi), Z_n(\phi))/\sqrt{n}$ are asymptotically independent Gaussian random variables with asymptotic variances equal to $(f(\theta), f(\theta), f(\phi), f(\phi))/2$, where $f(\cdot)$ is the spectral density function.*

### 3.2. Bias of the periodogram

It is known that for regularly observed stationary time series the periodogram is an asymptotically unbiased estimator of the spectral density function. This is not the case for irregular spatial data. In this section, we provide an expression for the bias of the periodogram.

For the process $(X_i)$ given in (2.1), assume that the mean is 0 and define the covariance function $\gamma_k = \mathbb{E}(X_0 X_k)$, $k \in \mathbb{Z}^d$. By (3.2), we have $\sum_{k\in\mathbb{Z}^d} |\gamma_k| < \infty$. Define the spectral density

$$f(\theta) = \sum_{k\in\mathbb{Z}^d} \gamma_k \cos(k'\theta). \tag{3.5}$$

In the literature the scaled form by $(2\pi)^{-d}$ is also widely used. In this paper we use the form (3.5). Recall that $\Gamma_n = \{L_1, \dots, L_n\}$. Let $J = \{k \in \mathbb{Z}^d : \exists\ i, j \text{ with } L_i - L_j = k\}$ and $m_k = \#\{(i,j) : L_i - L_j = k\}$. We define the *location adjusted spectral density function* by

$$f_J(\theta) = \mathbb{E}[I_n(\theta)] = \frac{1}{n}\sum_{j\in J} m_j \gamma_j \cos(j'\theta). \tag{3.6}$$

The bias of $I_n(\theta)$ is

$$B_n(\theta) = f(\theta) - \mathbb{E}[I_n(\theta)] = \sum_{k\in\mathbb{Z}^d}\left(1 - \frac{m_k}{n}\right)\gamma_k \cos(k'\theta). \tag{3.7}$$

If, for any fixed $k$, $m_k/n \to 1$ (which holds for the regular rectangle index set $\Gamma_n = \prod_{i=1}^d \{1, 2, \dots, J_i\}$, where $J_i \asymp n^{1/d}$ and $\prod_{i=1}^d J_i = n$), then by the Lebesgue dominated convergence theorem, $B_n(\theta) \to 0$ as $n \to \infty$. However, the same cannot be said for the irregular spatial case. In particular, if for each $k \in \mathbb{Z}^d$, the ratio $m_k/n$ approaches $r_k$ as $n \to \infty$ and these $r_k$'s are constant, then the asymptotic bias is given by $B(\theta) = \sum_{k\in\mathbb{Z}^d}(1 - r_k)\gamma_k \cos(k'\theta)$.

## 4. Whittle likelihood and parametric spectral estimate

In this section we shall discuss a parametric estimator of the spectral density function $f(\theta)$. In what follows, we write the spectral density $f(\theta)$ (where $\theta \in \mathbb{R}^d$) as $f_\alpha(\theta)$ for a certain parameter vector $\alpha \in \mathbb{R}^p$. Since the spectral density governs the covariance function of a stationary process, $\gamma_k$ will also be a function of $\alpha$. Therefore, the location adjusted spectral density function $f_J(\theta)$ and the bias $B_n(\theta)$ discussed in the previous section are functions of both $\theta$ and $\alpha$. We will denote them by $f_{J,\alpha}(\theta)$ and $B_{n,\alpha}(\theta)$, respectively.

A widely used approach to estimate the unknown parameter $\alpha$ is to minimize the (negative) Whittle's likelihood or an approximation to the Gaussian log likelihood which is of the form

$$p_n(\alpha) = \int_D \left\{\log f_\alpha(\theta) + \frac{I_n(\theta)}{f_\alpha(\theta)}\right\} d\theta, \text{ where } D = [-\pi, \pi]^d. \tag{4.1}$$

Dahlhaus and Künsch (1987) developed an asymptotic theory for Whittle estimator for regularly spaced time series data using a bias-adjusted periodogram in place of $I_n(\theta)$. We will adopt their technique to correct for the bias of the periodogram. The Whittle likelihood for the irregularly spaced data is

$$p_n(\alpha) = \int_D \left[\log f_{J,\alpha}(\theta) + \frac{I_n(\theta)}{f_{J,\alpha}(\theta)}\right] d\theta. \tag{4.2}$$

We denote the Whittle estimator that minimizes $p_n(\alpha)$ by $\hat{\alpha}_n$ and study the asymptotic behavior of this estimator. Suppose the parameter space for $\alpha$ is $A$. Let the true value of the parameter be $\alpha_0$, and suppose that the Whittle estimate $\hat{\alpha}_n$ exists in the parameter space $A$ for all $n$. Before stating the theorem for the Whittle estimator, in addition to the asymptotic regime (Assumption 1) and the nonlinear nature of the stationary random field (Assumption 2), we list a few more assumptions:

**Assumption 3.** *The parameter space $A \subset \mathbb{R}^p$ is compact, and $D \subset \mathbb{R}^d$ is symmetric and compact such that the spectral density function (now denoted as $f_{J,\alpha}(\theta)$, defined on $A \times D$) is twice differentiable with respect to $\alpha$ and the first and second order derivatives are continuous for $\theta \in D$.*

**Assumption 4.** *(Identifiability condition) For $\alpha_1 \neq \alpha_2$, $f_{J,\alpha_1}(\theta) \neq f_{J,\alpha_2}(\theta)$ on a subset of $D$ with positive Lebesgue measure.*

**Assumption 5.** *If $\nabla$ denotes the first order derivative of a function, then*

$$\int_D \frac{|\nabla f_{J,\alpha_0}(\theta)|}{\{f_{J,\alpha_0}(\theta)\}^2} d\theta < \infty \quad \textit{and} \quad \int_D \left(\frac{|\nabla f_{J,\alpha_0}(\theta)|}{f_{J,\alpha_0}(\theta)}\right)^2 d\theta < \infty.$$

Finally, let us use $p(\alpha)$ to denote the limit of the Whittle likelihood $p_n(\alpha)$ defined by (4.2). Note that $f_{J,\alpha}(\theta)$ and consequently, $p(\alpha)$ depends on the index set $\Gamma_n$. Also, for any $\Gamma_n$, we can say that $p(\alpha) > p(\alpha_0)$, for any $\alpha \neq \alpha_0$.

We now discuss the connection with the regularity assumptions. If the functional dependence measure $\delta_{i,p}$ satisfies the summability condition

$$\sum_{i \in \mathbb{Z}^d} |i|^2 \delta_{i,2} < \infty,$$

then the spectral density function is a bounded, twice partially differentiable function in view of

$$\sum_{i \in \mathbb{Z}^d} |i|^2 |\gamma_i| \leq \sum_{i \in \mathbb{Z}^d} |i|^2 \delta_{i,2} \sum_{i \in \mathbb{Z}^d} \delta_{i,2} < \infty.$$

Assumptions 3 to 5 are similar to the ones used in Matsuda and Yajima (2009). Theorem 4.1 below concerns the asymptotic properties of the Whittle estimator.

**Theorem 4.1.** *Let $f_{J,\alpha}(\theta)$ be the location adjusted spectral density function (3.6) and denote the true value of the parameter $\alpha$ by $\alpha_0$ and assume that the Whittle estimator $\hat{\alpha}_n$ exists in the parameter space $A$ for all large $n$. Then, under Assumptions 1-5, the estimator $\hat{\alpha}_n$ satisfies the following:*

(a) **(Consistency)** *$\hat{\alpha}_n \to \alpha_0$ in probability as $n \to \infty$.*

(b) **(Asymptotic normality)** *Suppose, $h_\alpha(\theta) = \nabla(f_{J,\alpha}(\theta))^{-1}$. Then, the Whittle estimator $\hat{\alpha}_n$ satisfies the following ($L(.,.)$ is the Levy distance):*

$$L\bigg(\frac{\sqrt{n}(\hat{\alpha}_n - \alpha_0)}{f_{J,\alpha_0}(0)}, N_p\big(0, 2\Gamma(\alpha_0)\Sigma_n^2\Gamma(\alpha_0)\big)\bigg) \to 0 \qquad \text{as} \;\; n \to \infty \tag{4.3}$$

*where $n\Sigma_n^2 = \sum_{j,k\in\Gamma_n} (\int_D \exp\{\imath(k-j)'\theta\}h_{\alpha_0}(\theta)d\theta)(\int_D \exp\{\imath(k-j)'\theta\}h_{\alpha_0}(\theta)d\theta)'$, and $\Gamma(\alpha)^{-1} = \int_D \nabla f_{J,\alpha}(\theta)\nabla(f_{J,\alpha}(\theta))^{-1}\, d\theta$.*

**Remark 1.** *If for any fixed $k$, $m_k/n \to 1$, then $f_{J,\alpha}(\theta)$ converges to $f_\alpha(\theta)$. Hence, in the regular setup, as $n \to \infty$,*

$$\sqrt{n}(\hat{\alpha}_n - \alpha_0) \Rightarrow N_p\big(0, 2f_\alpha^2(0)\Gamma(\alpha_0)\Sigma_n^2\Gamma(\alpha_0)\big).$$

From a practical point of view, it is not possible to use the above theorem directly to find a confidence set for the true value of the parameter $\alpha_0$, for the covariance matrix in the above theorem depends on $\alpha_0$. Next, we describe a subsampling procedure to find a confidence set of $\alpha$. However, for irregularly spaced stationary random field, a subsampling method will not work in most cases. Here, we describe a method only for a regular random field.

For the following discussion, let us assume that we have data $(X_i)_{i\in I}$ from a random field indexed by a rectangle $I$ in $\mathbb{Z}^d$. For convenience, since we are going to consider regular spaced data, let us assume that $I = \{1, 2, \dots, l\}^d$. And suppose the Whittle likelihood estimator for this data is $\hat{\alpha}_l$.

To use the subsampling procedure, we will be considering smaller blocks from $I$. For a point $t = (t_1, \dots, t_d)$, $\hat{\alpha}_{l,t,b}$ will denote the Whittle likelihood estimator based on the data $(X_i)_{i\in I_{t,b}}$ where $I_{t,b} = \prod_{i=1}^d \{t_i, \dots, t_i + b\}$. Naturally, these estimates can be obtained for all $t$ such that $I_{t,b} \subset I$. Let us use $Q_{l,b}$ to denote all such $t$'s.

We are going to show that an adjusted empirical distribution of the Whittle likelihood estimators for all possible blocks is essentially an approximation for the limiting distribution of $\hat{\alpha}_l$, described in the previous theorem. To do that, we will look at indicators of Borel sets. For any Borel set $A \in \mathbb{R}^p$, let us use $F(A)$ to denote the limiting value of $P[c_l(\hat{\alpha}_l - \alpha_0) \in A]$ where $c_l$ is an appropriate scaling constant (see Remark 1). Analogously, $F_b(A)$ denotes the same for the subsamples and so, $F_b(A) \to F(A)$, as $b \to \infty$. We are going to consider the following empirical distribution of the subsampled Whittle likelihood estimators:

$$L_{l,b}(A) = \frac{1}{|Q_{l,b}|} \sum_{t\in Q_{l,b}} \mathbb{I}\{c_b(\hat{\alpha}_{l,t,b} - \hat{\alpha}_l) \in A\}. \tag{4.4}$$

Then, the following theorem proves the consistency of the above approximation and helps us determine a confidence set for $\alpha_0$.

**Theorem 4.2.** *Assume that $b \to \infty, b/n \to 0$ as $n \to \infty$. Then, $L_{l,b}(A)$ in the above definition goes to $F(A)$ in probability, for each Borel set $A$ whose boundary has mass zero under $F(\cdot)$.*

## 5. Non-parametric estimator of spectral density function

In this section, we shall study non-parametric kernel spectral density estimators of $f$ and their large-sample properties such as consistency and asymptotic normality. For a symmetric kernel function $K(\cdot)$ and for bandwidth $B_n$ we define the kernel spectral density estimator

$$f_n(\theta) = \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} X_{L_j} X_{L_k} K\left(\frac{L_j - L_k}{B_n}\right) e^{\imath (L_j - L_k)'\theta}. \tag{5.1}$$

For the consistency of the above estimator, we choose the kernel $K$ to satisfy the following

**Condition 1.** *The kernel function $K$ is symmetric, has support $[-1, 1]$, $K(0) = 1$, and $\sup_{|x|<1} |K'(x)| < \infty$.*

The above condition readily implies that $\kappa = \int_{-\infty}^{\infty} K^2(x)dx < \infty$, a quantity that will be needed later. A simple choice that satisfies the above properties is the rectangular kernel $K(x) = \mathbb{I}_{\{|x|\leq 1\}}$. The following theorem asserts the consistency result of the nonparametric estimator above.

**Theorem 5.1.** *Assume that $\mathbb{E}(X_k) = 0$, $X_k \in L^p$, $p \geq 2$ and $\Theta_{0,p} = \sum_{j=0}^{\infty} \delta_{j,p} < \infty$, the bandwidth $B_n \to \infty$ and $B_n = o(n)$ as $n \to \infty$. Then, under Condition 1,*

$$\sup_{\theta \in \mathbb{R}^d} \|f_n(\theta) - \mathbb{E} f_n(\theta)\|_{p/2} \to 0.$$

**Corollary 5.1.** *Assume there exists a constant $c > 0$ such that $f_n(0) > c$. Let conditions in Theorem 5.1 be satisfied and let $\bar{X}_n$ be the mean of the sample $\{X_i, i \in \Gamma_n\}$. Then,*

$$\frac{\sqrt{n}\bar{X}_n}{\sqrt{f_n(0)}} \Rightarrow N(0, 1). \tag{5.2}$$

*Proof.* Let $v_n = n\mathbb{E}(\bar{X}_n^2)$. By the Lebesgue dominated convergence theorem,

$$v_n - \mathbb{E}\big(f_n(0)\big) = \frac{1}{n} \sum_{k \in \mathbb{Z}^d} m_k \big[1 - K(k/B_n)\big] \gamma_k \to 0,$$

where $m_k$ is as defined in Section 3.2. By Theorem 3.1, $\sqrt{n}\bar{X}_n/\sqrt{v_n} \Rightarrow N(0, 1)$. Hence by Theorem 5.1 and Slutsky's theorem, (5.2) follows. □

One can apply Corollary 5.1 to construct confidence intervals for the mean $\mu$ based on irregularly spaced data $X_{L_1}, \ldots, X_{L_n}$. Let $\tilde{f}_n(\theta)$ be defined as $f_n(\theta)$ in (5.1) with $X_j$ therein replaced by $X_j - \bar{X}_n$. Given $0 < \alpha < 1$, the $(1 - \alpha)$th confidence intervals for the mean $\mu$ is $\bar{X}_n \pm z_{1-\alpha/2}\sqrt{\tilde{f}_n(0)/n}$.

Next, we discuss the asymptotic distribution of the non-parametric spectral density estimator in (5.1). The proof of the theorems are given in the Appendix.

**Theorem 5.2.** *Let Condition 1 be satisfied and define* $\kappa = \int_{-\infty}^{\infty} K^2(x)dx < \infty$. *Assume* $\mathbb{E}(X_k) = 0$, $\mathbb{E}(X_k^4) < \infty$, $\Theta_{0,4} < \infty$, $B_n \to \infty$ *and* $B_n = o(n)$ *as* $n \to \infty$. *Then, for any fixed* $\theta \in \mathbb{R}^d$,

$$\sqrt{\frac{n}{B_n}}\left(\frac{f_n(\theta) - \mathbb{E}[f_n(\theta)]}{f_n(\theta)}\right) \Rightarrow N\big(0, \kappa\big). \tag{5.3}$$

**Remark 2.** *An interesting observation is that the variance of the estimate* $f_n(\theta)$ *is asymptotically equal to* $\mathbb{E}(f_n(\theta))^2$, *multiplied by an appropriate scaling term. Therefore, the concepts of variance stabilizing transformation and delta method tell us that we can take the logarithm of the estimate to obtain*

$$\sqrt{\frac{n}{B_n}}\Big(\log f_n(\theta) - \log \mathbb{E}[f_n(\theta)]\Big) \Rightarrow N\big(0, \kappa\big).$$

*This result can be used to form a confidence interval for* $\mathbb{E}(f_n(\theta))$ *of the form* $\exp(\log f_n(\theta) \pm z_{1-\alpha/2}\sqrt{\kappa B_n/n})$.

## 6. Estimation of covariance matrices

In spatial statistics, a fundamentally important problem is to estimate the covariance matrix of the data. It is useful in many aspects of multivariate analysis including principal component analysis, linear discriminant analysis and graphical modeling. One can infer dependence structures among variables by estimating the associated covariance matrices. In this section, we will discuss the estimation of the covariance matrix of an irregular spaced data $(X_{L_1}, \dots, X_{L_n})$. Now, to judge the quality of a matrix estimate, we will use the operator norm. For an estimate of the covariance matrix, we are going to use the $l_2$ "operator norm" and give an upper bound for this. Recall that $l_2$ norm or spectral radius of a matrix $A$ is defined as $\rho(A) = \max_{|x|=1} |Ax|$.

Let $\Sigma_n = (\gamma_{L_i - L_j})_{1\le i,j\le n}$ be the covariance matrix to be estimated. Its estimator $\hat{\Sigma}_n$ is defined by

$$\hat{\Sigma}_n = (\hat{\gamma}_{L_i - L_j})_{1\le i,j\le n} \qquad \text{where} \;\; \hat{\gamma}_k = \frac{1}{m_k}\sum_{i,j:L_i-L_j=k} (X_{L_i} - \bar{X})(X_{L_j} - \bar{X}).$$

In the above, $m_k = \#\{(i,j) : L_i - L_j = k\}$. Based on the known inconsistency results for the periodogram estimate, it can be shown that $\hat{\Sigma}_n$ is not a consistent estimate (Wu and Pourahmadi (2009)). So, instead of this, we will use non-parametric kernel-based estimators, similar to what was used for the spectral density. More precisely, we define the following estimate for the covariance matrix:

$$\hat{\Sigma}_{n,B_n} = \left(\hat{\gamma}_{L_i - L_j} K\left(\frac{L_i - L_j}{B_n}\right)\right)_{1\le i,j\le n}, \tag{6.1}$$

where $K$ is a kernel function and $B_n$ is a bandwidth sequence satisfying appropriate conditions, as discussed below. For this banded covariance matrix estimate, we prove the following

**Theorem 6.1.** *Suppose $\{X_t\}_{t\in\mathbb{Z}^d}$ is a stationary process and each $X_i \in L^p$ for some $p \in (2,4]$. If $m_k \asymp n$, the kernel $K$ satisfies Condition 1, the bandwidth $B_n$ satisfies the conditions $B_n \to \infty$ and $B_n/n^\delta \to 0$ for some $\delta > (1-2/p)/d$, then the estimator $\hat{\Sigma}_{n,B_n}$ in equation (6.1) is consistent and the spectral radius $\rho(\hat{\Sigma}_{n,B_n} - \mathbb{E}(\hat{\Sigma}_{n,B_n}))$ has a convergence rate of $O_P(B_n^d n^{2/p-1})$.*

**Remark 3.** *Note that the above covariance matrix estimate is not necessarily non-negative definite. If we define a matrix $K_{n,B_n} = (K((L_i - L_j)/B_n))_{1\le i,j\le n}$, then the covariance matrix estimate can be written as $\hat{\Sigma}_{n,B_n} = \hat{\Sigma}_n \star K_{n,B_n}$; where $\star$ is the Hadamard or Schur product, which is formed by the element-wise multiplication of matrices.*

*By Schur Product theorem, since $\hat{\Sigma}_n$ is already non-negative definite, the Schur product will be non-negative whenever $K_{n,B_n}$ is non-negative definite. One particular example is the triangular kernel $K(u) = \max(0, 1-|u|)$ which would lead to a positive definite weight matrix $K_{n,B_n}$. Thus, using this kernel function will give us a non-negative definite covariance matrix estimate $\hat{\Sigma}_{n,B_n}$.*

## 7. A simulation study

In this section, we assess empirically the impact of the sampling index set $\Gamma_n$ on the parameter estimation procedure. For that, we consider an isotropic spatial auto-regressive (AR) model in a two-dimensional setting. This model is similar to the isotropic spatial AR model discussed by Azomahou (2009) and Lavancier (2011). More precisely, in our spatial AR model, each observation is dependent on the four neighbors in the following way:

$$X_{i,j} = c(X_{i-1,j} + X_{i+1,j} + X_{i,j+1} + X_{i,j-1}) + \varepsilon_{i,j} + \varepsilon_{i,j-1}\varepsilon_{i,j+1}, \quad \text{for} \quad i,j \in \mathbb{Z}, \tag{7.1}$$

where the parameter $c$ gauges the strength of dependence of an observation on its four neighbors, and the innovations $\varepsilon_{i,j}$ are generated independently from a standard normal distribution. Note that this model is an example of non-linear random fields. For more detailed information on such models; see (Cressie, 2015, Chapter 6).

We start with a full grid of size $n$-by-$n$ from which we would like to generate sample data of different sizes. In order to generate the full data, we consider it in a vector form $\mathbf{X}$, such that the above model helps us write it as $\mathbf{X} = A\mathbf{X} + \mathbf{e}$, and thereby, $\mathbf{X} = (I - A)^{-1}\mathbf{e}$. Here, $A$ an $n^2 \times n^2$ sparse matrix (non-zero elements are the parameter of the model) and $\mathbf{e}$ is a vector such that each component is of the form $\varepsilon_{i,j} + \varepsilon_{i,j-1}\varepsilon_{i,j+1}$.

In our simulation study, we generate a data-set on a grid of $75 \times 75$. Then, we take different samples of varying sizes to estimate the parameter $c$ using the Whittle likelihood approach and calculate the mean squared error (MSE) of the estimates. For each of the sample sizes, we also choose a regular grid and compare the performances of the regular setup to that of the irregular setup. The parameters and the set-up of the simulation are described below:

- We used the parameter $c = 0.2$.
- The innovations $\varepsilon_{i,j}$ are generated independently from a $N(0,1)$ distribution.
- The grid consists of $75 \times 75$ data points and we take samples of different sizes (between $10^2$ and $40^2$). First, these samples are chosen randomly from the whole grid, to ensure that we have an irregularly spaced data. And then, we choose a regular grid of similar size to compare the performances.
- For each sample size, the experiment is repeated multiple times, and the mean squared error for the parameter estimate (error calculated from the actual value $c = 0.2$) is calculate for the regular and irregular data. For the irregular set-up, the Whittle likelihood is defined using the location adjusted spectral density, as shown in equation (4.2). In addition to that, we also perform same analysis for Whittle likelihood defined in the original way (4.1). The results are shown in Table 1 below.

TABLE 1
*Mean squared error in estimating parameter c using simulated data.*

| Sample Size | Regular | Irregular (location adjusted) | Irregular (original) |
|---|---|---|---|
| 100 | 0.0004 | 0.0015 | 0.0239 |
| 225 | 0.0004 | 0.0012 | 0.0158 |
| 400 | 0.0001 | 0.0009 | 0.0148 |
| 625 | 0.0004 | 0.0005 | 0.0117 |
| 900 | 0.0009 | 0.0003 | 0.0109 |
| 1225 | 0.0004 | 0.0002 | 0.0096 |
| 1600 | 0.0001 | 0.0001 | 0.0081 |

The results of the simulation confirm that the estimation performance in the irregular set-up is comparable to the regular set-up when the sample size is larger than $20^2$. However, the regular set-up outperforms the irregular one for smaller sample sizes. On the other hand, when we perform the simulation for the original definition of the Whittle likelihood, the results are not at par with the location adjusted version. In fact, the mean squared error is very big for smaller sample sizes, but it reduces steadily as we take bigger samples. For sample size $40^2$, it is about 0.0081, which clearly establishes that as more and more samples are taken, the location adjusted spectral density approaches the true value of the spectral density. This is expected, and follows what was discussed in Section 4.

## 8. Appendix

### 8.1. *Proofs of theorems*

**Proof of Theorem 3.1.** It is worth mention that this proof is somewhat similar to theorem 1 in El Machkouri et al. (2013). However, for completeness of our paper, we are giving a detailed proof here.

At first, assume that $\liminf_n v_n^2/n > 0$. Then, there exists a constant $c_0 > 0$ and $n_0 \in \mathbb{N}$ such that $n/v_n^2 \leq c_0$ for any $n \geq n_0$.

Let $(m_n)_{n\geq 1}$ be a sequence of positive integers going to infinity and denote $\tilde{X}_j = \mathbb{E}(X_j \,|\, \mathcal{F}_{m_n}(j))$ where $\mathcal{F}_{m_n}(j) = (\varepsilon_{j-s}; |s| \leq m_n)$. So, there exists a measurable function $h$ such that $\tilde{X}_j = h(\varepsilon_{j-s}; |s| \leq m_n)$. Similarly, for a coupled process (see Definition 2.1), we can write

$$\tilde{X}_j^* = h(\varepsilon_{j-s}^*; |s| \leq m_n) = \mathbb{E}\big(X_j^* \,|\, \mathcal{F}_{m_n}^*(j)\big),$$

where $\mathcal{F}_{m_n}^*(j) = (\varepsilon_{j-s}^*; |s| \leq m_n)$. Also, for any $j \in \mathbb{Z}^d$, define

$$\delta_{j,p}^{(m_n)} = \big\|(X_j - \tilde{X}_j) - (X_j - \tilde{X}_j)^*\big\|_p.$$

In order to prove the theorem, we will need the following results.

- Using lemma 8.1, denoting $\Delta_p^{(m_n)} = \sum_{j\in\mathbb{Z}^d} \delta_{j,p}^{(m_n)}$, for any $n \in \mathbb{N}$ and any $p \geq 2$,
$$\Big\|\sum_{i\in\Gamma_n} a_i(X_i - \tilde{X}_i)\Big\|_p \leq \Big(2p\sum_{i\in\Gamma_n} a_i^2\Big)^{1/2} \Delta_p^{(m_n)}. \tag{8.1}$$
- If $\Delta_p < \infty$, then for any fixed $p \geq 0$, $\Delta_p^{(m_n)} \to 0$ as $n \to \infty$. To this end, note that
$$\begin{aligned} \delta_{j,p}^{(m_n)} &\leq \big\|X_j - X_j^*\big\|_p + \big\|\tilde{X}_j - \tilde{X}_j^*\big\|_p \\ &= \delta_{j,p} + \big\|\mathbb{E}(X_j \,|\, \mathcal{F}_{m_n}(j), \mathcal{F}_{m_n}^*(j)) - \mathbb{E}(X_j^* \,|\, \mathcal{F}_{m_n}^*(j), \mathcal{F}_{m_n}(j))\big\| \\ &\leq 2\delta_{j,p}. \end{aligned}$$
Since $\lim_{n\to\infty} \delta_{j,p}^{(m_n)} = 0$ and $\sum_{j\in\mathbb{Z}^d} \delta_{j,p} = \Delta_p < \infty$, using the dominated convergence theorem, we can say that $\lim_{n\to\infty} \Delta_p^{(m_n)} = 0$.
- Define $\tilde{W}_n = \sum_{j\in\Gamma_n} c_j \tilde{X}_j$. Then, using the last two results, we can write
$$\begin{aligned} \frac{\big\|W_n - \tilde{W}_n\big\|_2}{v_n} &= \frac{1}{v_n}\Big\|\sum_{i\in\Gamma_n} c_i(X_i - \tilde{X}_i)\Big\|_2 \\ &\leq \frac{2\Delta_2^{(m_n)}}{v_n}\Big(\sum_{i\in\Gamma_n} c_i^2\Big)^{1/2} \leq \big(2\Delta_2^{(m_n)}\big)\Big(\frac{n}{v_n^2}\Big)^{1/2}. \end{aligned}$$
And hence, using the assumption mentioned earlier,
$$\limsup_{n\to\infty} \frac{\big\|W_n - \tilde{W}_n\big\|_2}{v_n} = 0. \tag{8.2}$$
- We will also use the following central limit theorem, due to Heinrich (1988).

**Theorem 8.1.** *Let $(\Gamma_n)_{n\geq 1}$ be a sequence of finite subsets of $Z^d$ with $|\Gamma_n| \to \infty$ as $n \to \infty$ and $(m_n)_{n\geq 1}$ be a sequence of positive integers. For each $n \geq 1$, let $\{U_n(j), j \in \mathbb{Z}^d\}$ be an $m_n$-dependent random field with*

*$\mathbb{E}(U_n(j)) = 0$ for all $j \in \mathbb{Z}^d$. Also, assume that $\mathbb{E}(\sum_{j\in\mathbb{Z}^d} U_n(j))^2 \to \sigma^2$ as $n \to \infty$ where $\sigma^2$ is finite. Then, $\sum_{j\in\mathbb{Z}^d} U_n(j)$ converges in distribution to $N(0,\sigma^2)$ if there exists a finite constant $c > 0$ such that for any $n \geq 1$, $\sum_{j\in\mathbb{Z}^d} \mathbb{E}(U_n^2(j)) \leq c$ and for any $\epsilon > 0$ it holds that*

$$\lim_{n\to\infty} L_n(\epsilon) := m_n^{2d} \sum_{j\in\mathbb{Z}^d} \mathbb{E}\Big(U_n^2(j)\mathbb{I}\{|U_n(j)| \geq \epsilon m_n^{-2d}\}\Big) = 0. \tag{8.3}$$

We now apply the above results to prove our theorem. At first, define $U_n(j) := c_j\tilde{X}_j/v_n$, where $c_j$ is same as the coefficient of $X_j$ in $W_n$. Note that this definition satisfies the criteria that $\{U_n(j), j \in \mathbb{Z}^d\}$ is an $m_n$-dependent random field with $\mathbb{E}(U_n(j)) = 0$ for all $j \in \mathbb{Z}^d$. Further,

$$\mathbb{E}\Big(\sum_{j\in\mathbb{Z}^d} U_n(j)\Big)^2 = \frac{1}{v_n^2}\mathbb{E}\Big(\sum_{j\in\mathbb{Z}^d} c_j\tilde{X}_j\Big)^2 = \frac{\mathbb{E}\big(\tilde{W}_n^2 - v_n^2\big)}{v_n^2} + 1.$$

Now,

$$\begin{aligned}
\Big|\mathbb{E}\big(\tilde{W}_n^2 - v_n^2\big)\Big| &= \Big|\mathbb{E}\big(\tilde{W}_n^2 - W_n^2\big)\Big| \\
&\leq \big\|\tilde{W}_n + W_n\big\|_2\big\|\tilde{W}_n - W_n\big\|_2 \\
&\leq \big\|\tilde{W}_n - W_n\big\|_2\big(\big\|\tilde{W}_n - W_n\big\|_2 + 2\|W_n\|_2\big) \\
&\leq 2\Delta_2^{(m_n)}\big(\sum_{i\in\Gamma_n} c_i^2\big)^{1/2}\big(2\Delta_2^{(m_n)}\big(\sum_{i\in\Gamma_n} c_i^2\big)^{1/2} + 4\Delta_2\big(\sum_{i\in\Gamma_n} c_i^2\big)^{1/2}\big) \\
&= 4\Delta_2^{(m_n)}\big(\sum_{i\in\Gamma_n} c_i^2\big)\big(\Delta_2^{(m_n)} + 2\Delta_2\big)
\end{aligned}$$

Thus, using our assumptions, $|\mathbb{E}(\tilde{W}_n^2 - v_n^2)/v_n^2| \leq 4\Delta_2^{(m_n)}\big(\Delta_2^{(m_n)} + 2\Delta_2\big)(n/v_n^2) \to 0$. And hence, as $n \to \infty$, $\mathbb{E}\big(\sum_{j\in\mathbb{Z}^d} U_n(j)\big)^2 \to 1$. On the other hand, for any $n \geq n_0$,

$$\sum_{j\in\mathbb{Z}^d} \mathbb{E}(U_n^2(j)) = \frac{1}{v_n^2}\sum_{j\in\mathbb{Z}^d} c_j^2\mathbb{E}(\tilde{X}_j^2) \leq \frac{n}{v_n^2}\mathbb{E}(\tilde{X}_0^2) \leq c_0\mathbb{E}(\tilde{X}_0^2).$$

Finally, let $\epsilon > 0$ be fixed. Since $|c_j| \leq 1$, we have

$$\mathbb{I}\Big\{|U_n(j)| \geq \epsilon m_n^{-2d}\Big\} = \mathbb{I}\Big\{|\tilde{X}_j| \geq \frac{\epsilon v_n}{|c_j|}m_n^{-2d}\Big\} \leq \mathbb{I}\Big\{|\tilde{X}_j| \geq \frac{\epsilon v_n}{m_n^{2d}}\Big\},$$

and then, we get

$$\begin{aligned}
L_n(\epsilon) &\leq \frac{m_n^{2d}}{v_n^2}\sum_{j\in\mathbb{Z}^d}\mathbb{E}\Big(\tilde{X}_j^2\mathbb{I}\Big\{|\tilde{X}_j| \geq \frac{\epsilon v_n}{m_n^{2d}}\Big\}\Big) \\
&\leq c_0 m_n^{2d}\mathbb{E}\Big(\tilde{X}_0^2\mathbb{I}\Big\{|\tilde{X}_0| \geq \frac{\epsilon v_n}{m_n^{2d}}\Big\}\Big)
\end{aligned}$$

$$\leq \quad c_0 m_n^{2d} \times \left[ v_n P\left( |\tilde{X}_0| \geq \frac{\epsilon v_n}{m_n^{2d}} \right) + \mathbb{E}\left( X_0^2 \mathbb{I}\left\{ |X_0| \geq \sqrt{v_n} \right\} \right) \right]$$

$$\leq \quad \frac{c_0 m_n^{6d} \mathbb{E}(X_0^2)}{\varepsilon^2 v_n} + c_0 m_n^{2d} \psi(\sqrt{v_n}), \quad \text{where } \psi(x) = \mathbb{E}\big(X_0^2 \mathbb{I}\{|X_0| \geq x\}\big).$$

In order to ensure that $L_n(\epsilon) \to 0$, define the sequence $(m_n)_{n\geq 1}$ by

$$m_n = \begin{cases} \min\left\{ \left[ \psi\big(\sqrt{v_n}\big)^{-\frac{1}{4d}} \right], \left[ v_n^{\frac{1}{12d}} \right] \right\} & \text{if } \psi\big(\sqrt{v_n}\big) \neq 0 \\ \left[ v_n^{\frac{1}{12d}} \right] & \text{if } \psi\big(\sqrt{v_n}\big) = 0 \end{cases}$$

where [.] is the greatest integer function. Since $v_n \to \infty$ and $\psi\big(\sqrt{v_n}\big) \to 0$, it is easy to observe that

$$m_n \to \infty, \quad \frac{m_n^{6d}}{v_n} \leq \frac{1}{\sqrt{v_n}} \to 0 \qquad \text{and} \qquad m_n^{2d} \psi(\sqrt{v_n}) \leq \sqrt{\psi(\sqrt{v_n})} \to 0.$$

Hence, applying Theorem 8.1 and using (8.2), we derive that

$$\frac{\tilde{W}_n}{v_n} \xrightarrow[n\to\infty]{L} N(0,1) \qquad \text{which implies that} \qquad \frac{W_n}{v_n} \xrightarrow[n\to\infty]{L} N(0,1).$$

So, we have proved required result (3.3) assuming that $\liminf_n v_n^2/n > 0$. If this condition fails, we can get a subsequence $n' \to \infty$ such that

$$L\left[ W_{n'}/\sqrt{|\Gamma_{n'}|}, N\big(0, v_{n'}^2/|\Gamma_{n'}|\big) \right] \to l, \text{ as } n' \to \infty, \tag{8.4}$$

for some $l \in [0, \infty]$. Furthermore, if $v_{n'}^2/|\Gamma_{n'}|$ does not converge to 0, we can get a further subsequence $n''$ such that $\liminf_{n''}(v_{n''}^2/|\Gamma_{n''}|) > 0$, implying

$$L\left[ W_{n''}/\sqrt{|\Gamma_{n''}|}, N\big(0, v_{n''}^2/|\Gamma_{n''}|\big) \right] \to 0, \text{ as } n'' \to \infty. \tag{8.5}$$

The above can be shown using the method we adopted previously and this contradicts (8.4). Consequently, we can say that $v_{n'}^2/|\Gamma_{n'}|$ converges to 0 and thus, $W_{n'}/\sqrt{|\Gamma_{n'}|}$ converges to 0 in probability, implying that the Levy distance between $W_{n'}/\sqrt{|\Gamma_{n'}|}$ and $N(0, v_{n'}^2/|\Gamma_{n'}|)]$ goes to 0, again contradicting (8.4). So, finally, proof of the first part of the theorem 3.1 is complete. The second part is just a corollary that follows easily from the first part. □

**Proof of Proposition 1.** Observe that $\mathbb{E}[S_n(\theta)S_n(\phi)] = \sum_{k\in\mathbb{Z}^d} \gamma_k a_{n,k}$, where the coefficient $a_{n,k}$ is equal to $n^{-1}\sum_{j,l\in\Gamma_n:j-l=k} e^{-\imath(j'\theta+l'\phi)}$. Clearly $|a_{n,k}| \leq 1$. Under the condition $\theta \neq \phi$, $\theta, \phi \in (-\pi, \pi]^d$, $\theta + \phi \neq 0$, we have for any fixed $k$, $\lim_{n\to\infty} a_{n,k} = 0$. By (3.2), we have $\sum_{k\in\mathbb{Z}^d} |\gamma_k| < \infty$, and by Lebesgue dominated convergence theorem, $\mathbb{E}[S_n(\theta)S_n(\phi)] \to 0$ as $n \to \infty$. Similarly,

$\mathbb{E}[S_n(\theta)S_n(-\phi)] \to 0$. Note that the conjugate of $S_n(\theta)$ is $S_n(-\theta)$. Hence the covariance matrix of $(Y_n(\theta), Z_n(\theta), Y_n(\phi), Z_n(\phi))/\sqrt{n}$ is asymptotically diagonal. That the diagonal elements are equal to $f(\cdot)/2$ can be obtained by noting that $\mathbb{E}[|S_n(\theta)|^2] \to f(\theta)$ and $\mathbb{E}[Y_n(\theta)^2 - Z_n(\theta)^2] \to 0$. The proposition then follows from Theorem 3.1. □

**Proof of Theorem 4.1, part (a).** Define

$$p(\alpha) := \int_D \left[\log\{f_{J,\alpha}(\theta)\} + \frac{f_{J,\alpha_0}(\theta)}{f_{J,\alpha}(\theta)}\right] d\theta.$$

So, for any $\alpha \neq \alpha_0$, we can get

$$p(\alpha) - p(\alpha_0) = \int_D \left[\log \frac{f_{J,\alpha}(\theta)}{f_{J,\alpha_0}(\theta)} + \frac{f_{J,\alpha_0}(\theta)}{f_{J,\alpha}(\theta)} - 1\right] d\theta > 0. \tag{8.6}$$

because the integrand is 0 for $\alpha = \alpha_0$ and otherwise always positive. Observe that $p_n(\alpha) \to p(\alpha)$ in probability, in view of Lemma 8.4 and the assumptions mentioned in Section 4. Thus, there exists a positive constant $C_\alpha$ such that

$$\lim_{n\to\infty} P\big[p_n(\alpha_0) - p_n(\alpha) < -C_\alpha\big] = 1. \tag{8.7}$$

Now, consider any two $\alpha_1, \alpha_2$ such that $\|\alpha_1 - \alpha_2\| < \delta$ for some small positive constant $\delta$, fixed a priori. Then, using the continuity of the functions, we can get that

$$\begin{aligned} |p_n(\alpha_1) - p_n(\alpha_2)| &\leq \left|\int_D \log \frac{f_{J,\alpha_1}(\theta)}{f_{J,\alpha_2}(\theta)}\, d\theta\right| + \int_D I_n(\theta)\left|\frac{1}{f_{J,\alpha_1}(\theta)} - \frac{1}{f_{J,\alpha_2}(\theta)}\right| d\theta \\ &\leq \delta\left(K_1 + K_2 \int_D I_n(\theta) d\theta\right), \end{aligned} \tag{8.8}$$

where $K_1, K_2$ are constants. Let us denote the above bound by $K_n(\delta)$ and observe that one can always find $\delta$ such that

$$\lim_{n\to\infty} P\big[K_n(\delta) < C_\alpha\big] = 1. \tag{8.9}$$

Now, for any particular $\alpha_1$, consider all points $\alpha_2$ such that $\|\alpha_1 - \alpha_2\| < \delta$. Then, using (8.8) and (8.9), we can say that for large $n$, $p_n(\alpha_1) - p_n(\alpha_2) \leq C_\alpha$ with probability going to 1. Combining this with (8.7) for $\alpha = \alpha_1$, we get that for any $\alpha_1$

$$\lim_{n\to\infty} P\left[\sup_{\alpha_2 : \|\alpha_1 - \alpha_2\| < \delta} \{p_n(\alpha_0) - p_n(\alpha_2)\} < 0\right] = 1. \tag{8.10}$$

For an $\alpha_1$, let us denote the above sets of the form $\{\alpha_2 : \|\alpha_1 - \alpha_2\| < \delta\}$ by $S(\alpha_1)$. To remain consistent with out notation, for $\alpha_0$, let us consider $S(\alpha_0) = \{\alpha_2 : \|\alpha_0 - \alpha_2\| < \gamma\}$. Observe that the above result is true for any possible value of $\gamma$. Now, for the whole parameter space $A$, if we consider the collection of subsets $\{S(\alpha) : \alpha \in A\}$, this forms an open cover of the whole parameter space and since $A$ is compact, it will have a finite cover. Let us denote this as $\{S(\alpha_i) : i = 0, 1, 2, \ldots, m \text{ and } \alpha_i \in A \ \forall\ i\}$. Note that from (8.10),

$$\lim_{n\to\infty} P\left[\sup_{\alpha\in\cup_{i=0}^m S(\alpha_i)} \{p_n(\alpha_0) - p_n(\alpha)\} < 0\right] = 1. \tag{8.11}$$

Now, $A = \cup_{i=0}^m S(\alpha_i)$ and from the above equation, we get that as $n \to \infty$, $p_n(\alpha_0) < \inf_A p_n(\alpha)$ with probability going to 1. Thus,

$$\lim_{n\to\infty} P\left[\inf_{\alpha\in S(\alpha_0)} p_n(\alpha) = \inf_{\alpha\in A} p_n(\alpha)\right] = 1. \tag{8.12}$$

Therefore, clearly, as $n \to \infty$, the minimizer will always be in $S(\alpha_0)$ and since we can fix $\gamma$ as small as possible, we can say that $\lim_{n\to\infty} P\big[|\hat{\alpha}_n - \alpha_0| < \gamma\big] = 1$ and so, $\hat{\alpha}_n \xrightarrow{P} \alpha_0$. Hence, the Whittle likelihood estimate is consistent. □

**Proof of Theorem 4.1, part (b).** Since $\hat{\alpha}_n$ is the minimizer for the Whittle likelihood, we will start with the Taylor series expansion and we get

$$\hat{\alpha}_n - \alpha_0 = -\left(\frac{\partial^2 p_n(\alpha^*)}{\partial\alpha\ \partial\alpha'}\right)^{-1} \frac{\partial p_n(\alpha_0)}{\partial\alpha}. \tag{8.13}$$

Now, we would consider the two terms on the right hand side separately. For the second term, using the assumptions, we have,

$$\begin{aligned}\frac{\partial p_n(\alpha_0)}{\partial\alpha} &= \int_D \left[\frac{1}{f_{J,\alpha_0}(\theta)} - \frac{I_n(\theta)}{\{f_{J,\alpha_0}(\theta)\}^2}\right] \frac{\partial}{\partial\alpha}\{f_{J,\alpha_0}(\theta)\}\ d\theta \\ &= \int_D \{I_n(\theta) - E(I_n(\theta))\} \frac{\partial}{\partial\alpha}\{f_{J,\alpha_0}(\theta)\}^{-1}\ d\theta\end{aligned}$$

Let us use $\nabla g$ and $\nabla^2 g$ to denote the first order derivative and the Hessian of a function $g$. Also, let $h_{\alpha_0}(\theta) = \nabla\{f_{J,\alpha_0}(\theta)\}^{-1}$. Note that this is non-stochastic, but depends on $n$ and the true value of $\alpha$. In order to find the asymptotic distribution of the above term, we will make use of Lemma 8.3. Observe that $I_n(\theta)$ can be written as

$$nI_n(\theta) = n\,|S_n(\theta)|^2 = \sum_{j,k\in\Gamma_n} X_j X_k \exp\{\imath(k-j)'\theta\}.$$

Now, if we consider the stochastic term in the integral above, we can write it as

$$\begin{aligned}nP_n = n\int_D I_n(\theta) h_{\alpha_0}(\theta)\ d\theta &= \sum_{j,k\in\Gamma_n} X_j X_k \int_D \exp\{\imath(k-j)'\theta\} h_{\alpha_0}(\theta)\ d\theta \\ &= \sum_{j,k\in\Gamma_n} X_j X_k \beta_{n,j-k}.\end{aligned}$$

Let us consider the asymptotic distribution of $c'P_n$ for some real-valued vector $c \in \mathbb{R}^p$. Note that the coefficients $c'\beta_{n,j-k}$ do not depend on $\theta$. Hence, comparing the above expression with that of Lemma 8.3, we can set $b_{n,j} = a_{n,j} = c'\beta_{n,j}$ and fix $\theta = 0$ in that theorem. Then,

$$\begin{aligned}\sigma_n^2 &= 2 \sum_{j,k\in\Gamma_n} \Big(\int_D \exp\{\imath(k-j)'\theta\} c' h_{\alpha_0}(\theta)\ d\theta\Big)^2 \\ &= 2c'\Bigg[\sum_{j,k\in\Gamma_n} \Big(\int_D \exp\{\imath(k-j)'\theta\} h_{\alpha_0}(\theta) d\theta\Big)\Big(\int_D \exp\{\imath(k-j)'\theta\} h_{\alpha_0}(\theta) d\theta\Big)'\Bigg] c.\end{aligned}$$

Based on the assumptions we have, we can say that the integrals are finite and on the other hand, it depends only on $\alpha_0$ (through $h(.)$). Let us now denote the matrix in the middle of the above expression by $n\Sigma_n^2$. Using the assumptions in Section 4, one can now easily check that (8.27) and other conditions, required to apply the lemma, are satisfied. So, using it, we get that

$$L\Bigg(nc'P_n(\theta) - nE[c'P_n(\theta)], N\big(0, f_{J,\alpha_0}^2(0)\cdot 2nc'\Sigma_n^2 c\big)\Bigg) \xrightarrow{n\to\infty} 0. \tag{8.14}$$

where $L(\cdot,\cdot)$ denotes the Levy distance. Since the above is true for all $c$, we can say that $L\big(nP_n(\theta) - n\mathbb{E}[P_n(\theta)], N(0, f_{J,\alpha_0}^2(0)\cdot 2n\Sigma_n^2)\big) \to 0$. Combining this with the expression obtained for $\partial p_n(\alpha_0)/\partial\alpha$, we obtain

$$L\Bigg(\sqrt{n}\bigg(\frac{\partial p_n(\alpha_0)}{\partial\alpha}\bigg), N\bigg(0, 2f_{J,\alpha_0}^2(0)\Sigma_n^2\bigg)\Bigg) \xrightarrow{n\to\infty} 0. \tag{8.15}$$

Now, for the first term, we need to take the second order derivative of the integrand in the expression of Whittle likelihood. Then,

$$\frac{\partial^2 p_n(\alpha^*)}{\partial\alpha\ \partial\alpha'} = \int_D \{I_n(\theta) - f_{J,\alpha^*}(\theta)\}\nabla^2\{f_{J,\alpha^*}(\theta)\}^{-1}\ d\theta - \Gamma(\alpha^*)^{-1}.$$

We already have proved that $\hat{\alpha}_n$ is consistent for $\alpha_0$. Using this, along with the assumptions mentioned in Section 4 and Lemma 8.4, the first term above goes to 0 as $n\to\infty$ and hence,

$$-\frac{\partial^2 p_n(\alpha^*)}{\partial\alpha\ \partial\alpha'} \to \int_D \nabla f_{J,\alpha_0}(\theta)\nabla\{f_{J,\alpha_0}(\theta)\}^{-1}\ d\theta = \Gamma(\alpha_0)^{-1}. \tag{8.16}$$

Finally, the central limit theorem follows from (8.13), (8.15) and (8.16). □

**Proof of Theorem 4.2.** We are going to set some notation at first.

For a Borel set $A$, let us use $\delta(A)$ to denote its boundary. Now, for a positive constant $\epsilon$, set $M_{A,\epsilon} = \cup_{x\in\delta(A)} B(x,\epsilon)$, where $B(x,\epsilon)$ denotes the closed ball with center $x$ and radius $\epsilon$. Let $A_{+\epsilon} = A \cup M_{A,\epsilon}$, $A_{-\epsilon} = A \cap M_{A,\epsilon}^c$. Thus, if $A$ is a ball with center $z$ and radius $r$, then $A_{+\epsilon}$ denotes the closed ball with center $z$ and radius $r+\epsilon$ while $A_{-\epsilon}$ denotes the open ball with center $z$ and radius $r-\epsilon$.

Let us also define the following empirical distribution for $\hat{\alpha}_{l,t,b}$:

$$L^0_{l,b}(A) = \frac{1}{|Q_{l,b}|} \sum_{t \in Q_{l,b}} \mathbb{I}\{c_b(\hat{\alpha}_{l,t,b} - \alpha_0) \in A\}. \tag{8.17}$$

And suppose $E_{l,b,\epsilon}$ denotes the event $\{\|c_b(\hat{\alpha}_l - \alpha_0)\| \leq \epsilon\}$. Because of the assumptions on $b$, we can easily say that $P(E_{l,b,\epsilon}) \to 1$ as $n \to \infty$, for any $\epsilon > 0$. Further note that

$$\begin{aligned} \mathbb{I}\{c_b(\hat{\alpha}_{l,t,b} - \alpha_0) \in A_{-\epsilon}\} \cdot \mathbb{I}\{E_{l,b,\epsilon}\} &\leq \mathbb{I}\{c_b(\hat{\alpha}_{l,t,b} - \hat{\alpha}_l) \in A\} \cdot \mathbb{I}\{E_{l,b,\epsilon}\} \\ &\leq \mathbb{I}\{c_b(\hat{\alpha}_{l,t,b} - \alpha_0) \in A_{+\epsilon}\}, \end{aligned}$$

and hence, with probability tending to one,

$$L^0_{l,b}(A_{-\epsilon}) \leq L_{l,b}(A) \leq L^0_{l,b}(A_{+\epsilon}).$$

Now, if we can prove that $L^0_{l,b}(A) \to F(A)$ for any Borel set $A$ whose boundaries have measure zero under $F(\cdot)$, then we would get that $F(A_{-\epsilon}) - \epsilon \leq L_{l,b}(A) \leq F(A_{+\epsilon}) + \epsilon$. Letting $\epsilon \to 0$ such that $A_{\pm\epsilon}$ are Borel sets whose boundaries have measure zero under $F(\cdot)$, we can then get that $L_{l,b}(A)$ is a good approximation for $F(A)$.

In order to show that $L^0_{l,b}(A) \to F(A)$, at first note that $\mathbb{E}[L^0_{l,b}(A)] = F_b(A) \to F(A)$ and thus, we only have to show that $\mathrm{Var}(L^0_{l,b}(A)) \to 0$ as $n \to \infty$.

We will be using $m$-dependence approximation to prove the required result. Here, we deal with the case $d = 2$ for convenience, but the proof would hold for any dimension. Let us write $(i_1, i_2)$ or $(j_1, j_2)$ for the 2-dimensional indexes $i, j$. Define the sigma field $\mathcal{F}_{i_1-m,i1} = \sigma(\varepsilon_j : j \in \mathbb{R}^d, i_1 - m \leq j_1 \leq i_1)$ and write $\mathcal{F}_{i_1}$ for $\mathcal{F}_{-\infty,i_1}$. The $m$-dependence approximation, for $m \geq 0$, is then defined by the following:

$$\tilde{X}_i := \mathbb{E}(X_i \,|\, \mathcal{F}_{i_1-m,i_1}). \tag{8.18}$$

Now, let us define a new quantity, as follows:

$$\tilde{L}_{l,b}(A) = \frac{1}{|Q_{l,b}|} \sum_{t \in Q_{l,b}} \mathbb{I}\{c_b(\tilde{\alpha}_{l,t,b} - \alpha_0) \in A\}. \tag{8.19}$$

Here, $\tilde{\alpha}_{l,t,b}$ is the Whittle likelihood estimator based on $\tilde{X}_t$'s. Below, for notational convenience, let us denote $\mathbb{I}\{c_b(\hat{\alpha}_{l,t,b} - \alpha_0) \in A\}$ by $Z_{t,b}$, $|Q_{l,b}|$ by $q$ and $\mathrm{Cov}(Z_{t,b}, Z_{t+k,b})$ by $\tau_k$. $\tilde{Z}_{t,b}$ and $\tilde{\tau}_k$ are defined analogously for $\tilde{X}_t$. Now, let $R_{s,m} = \{t \in Q_{l,b} : \left|\tau^{-1}(s) - \tau^{-1}(t)\right| \leq m\}$ and $R^c_{s,m} = Q_{l,b} - R_{s,m}$. Then,

$$\begin{aligned} \mathrm{Var}(\tilde{L}_{l,b}(A)) &= \frac{1}{q^2} \sum_{s \in Q_{l,b}} \sum_{t \in Q_{l,b}} \tilde{\tau}_{s-t} \\ &= \frac{1}{q^2} \sum_{s \in Q_{l,b}} \sum_{t \in R_{s,m}} \tilde{\tau}_{s-t} + \frac{1}{q^2} \sum_{s \in Q_{l,b}} \sum_{t \in R^c_{s,m}} \tilde{\tau}_{s-t} = A_1 + A_2 \quad \text{(say)} \end{aligned}$$

It is easy to note that that $A_1 = O(m^{1/d}q^{-1})$ and so, it goes to 0 as $n \to \infty$, for any fixed $m$. On the other hand, $A_2$ is exactly equal to 0, since for any $s$, $\tilde{\tau}_{s-t} = 0$ for all $t \in R^c_{s,m}$. The theorem is then proved in view of the fact that $\mathrm{Var}(L^0_{n,b}(A)) = \mathrm{Var}(\tilde{L}_{n,b}(A)) + o(1)$. □

**Proof of Theorem 5.1.** This proof will revolve around the notion of $m$-dependent processes, as defined by (8.18). Observe that, in this theorem, we are dealing with the quantity

$$\begin{aligned} f_n(\theta) &= \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n} X_{L_s}X_{L_t}K\left(\frac{L_s - L_t}{B_n}\right)e^{\imath(L_s-L_t)'\theta} \\ &= \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n} a_{n,L_s-L_t}X_{L_s}X_{L_t}, \qquad \text{where } \ a_{n,r} = K(r/B_n)e^{\imath r'\theta}. \end{aligned}$$

Based on the assumptions, it is easy to observe that $\sum_{r\in\mathbb{Z}}|a_{n,r}|^2 = O(B_n)$.

Now, since we intend to approximate using the $m$-dependent process, we should consider the following term, which is defined in a similar way as above, but with $\tilde{X}_{L_t}$'s. So, define $Y_t = \tilde{X}_{L_t}\sum_{s=1}^{t-k} a_{n,L_s-L_t}\tilde{X}_{L_s}$. Then

$$\begin{aligned} \tilde{f}_n(\theta) &= \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n} a_{n,L_s-L_t}\tilde{X}_{L_s}\tilde{X}_{L_t} \\ &= \frac{1}{n}\sum_{t=1}^{n}\tilde{X}^2_{L_t} + \frac{2}{n}\sum_{t=2}^{n}\tilde{X}_{L_t}\sum_{s=\max\{1,(t-k)\}}^{t-1} a_{n,L_s-L_t}\tilde{X}_{L_s} + \frac{2}{n}\sum_{t=k+1}^{n} Y_t. \end{aligned} \tag{8.20}$$

The idea here is to prove the required result in three steps. At first, we consider the last term in the above expression. Now, observe that the sequence $\{Y_{t+rk}\}_{r\geq 0}$ are $\mathrm{Ł}^p$ martingale differences whenever $|k| > m$. For convenience, let us take $|k| = 2m$. On the other hand, using Lemma 8.2, we get that

$$\begin{aligned} \|Y_t\|_p &= \|\tilde{X}_{L_t}\|_p\left\|\sum_{s=1}^{t-k} a_{n,L_s-L_t}\tilde{X}_{L_s}\right\|_p \\ &\leq \|X_0\|_p C_p\Theta_{0,p}\left(\sum_{s=1}^{t-k}|a_{n,L_s-L_t}|^2\right)^{1/2} = O(B_n^{1/2}). \end{aligned}$$

Now, we will write the last term in (8.20) using the sequences of the martingale differences and then it would satisfy the following. (Here, $N_j$ is used to denote the maximum possible index in that sequence and it will be of the order $n$.)

$$\left\|\frac{2}{n}\sum_{t=k+1}^{n} Y_t\right\|_p \leq \frac{2}{n}\sum_{j=1}^{k}\left\|\sum_{i=1}^{N_j} Y_{j+ik}\right\|_p = \frac{2}{n}\sum_{j=1}^{k} n^{1/2}O(B_n^{1/2}) = O[(mB_n/n)^{1/2}].$$

Let $\tilde{\gamma}_k = \mathbb{E}(\tilde{X}_0 \tilde{X}_k)$ and denote the last term in (8.20) by $W_n/n$. We are now going to consider the first two terms together in view of the fact that for any $l$, $\|n^{-1} \sum_t \tilde{X}_t \tilde{X}_{t+l} - \tilde{\gamma}_l\|_{p/2} \to 0$ as $n \to \infty$. Observe that in the first two terms, we are essentially combining all the terms of the form $\tilde{X}_{L_t} \tilde{X}_{L_s}$ where $|s-t| \leq k$. So, we can say that the following holds:

$$\left\|\tilde{f}_n(\theta) - W_n/n - \mathbb{E}[\tilde{f}_n(\theta) - W_n/n]\right\|_{p/2} \to 0.$$

Combining this with the above result that $\|W_n/n\|_p = O[(mB_n/n)^{1/2}]$, based on our assumption that $B_n/n \to 0$ as $n \to \infty$, we can conclude that

$$\|\tilde{f}_n(\theta) - \mathbb{E}[\tilde{f}_n(\theta)]\|_{p/2} \to 0 \qquad \text{as} \;\; n \to \infty. \tag{8.21}$$

Now, define the following:

$$S_n(\theta) = \sum_{j=1}^{n} X_{L_j} e^{\imath L_j' \theta} \qquad \text{and} \qquad \tilde{S}_n(\theta) = \sum_{j=1}^{n} \tilde{X}_{L_j} e^{\imath L_j' \theta}.$$

So, from the assumptions and using Lemma 8.2, we can say that $\|S_n(\theta) - \tilde{S}_n(\theta)\|_p = \Theta_{m,p} O(n^{1/2})$ and $\|S_n(\theta)\|_p + \|\tilde{S}_n(\theta)\|_p = O(n^{1/2})$.

Now, if $\hat{K}$ is the Fourier transform of $K$, we can write $f_n(\theta)$ with the help of $S_n$ in the form $nf_n(\theta) = \int \hat{K}(u)|S_n(B_n^{-1}u + \theta)|^2 \, du$ and we can use a similar definition for $\tilde{f}_n(\theta)$ using $\tilde{S}_n$. And then, the following can be obtained.

$$\begin{aligned} \|f_n(\theta) - \tilde{f}_n(\theta)\|_{p/2} &\leq \frac{1}{n} \int_{\mathbb{R}} |\hat{K}(u)| \left\| |S_n(B_n^{-1}u + \theta)|^2 - |\tilde{S}_n(B_n^{-1}u + \theta)|^2 \right\|_{p/2} du \\ &\leq \frac{1}{n} \int_{\mathbb{R}} |\hat{K}(u)| O(n) \Theta_{m,p} \, du = O(1)\Theta_{m,p}. \end{aligned} \tag{8.22}$$

Finally, observe that $\|f_n(\theta) - \mathbb{E}[f_n(\theta)]\|_{p/2} \leq \|f_n(\theta) - \tilde{f}_n(\theta)\|_{p/2} + \|\tilde{f}_n(\theta) - \mathbb{E}[\tilde{f}_n(\theta)]\|_{p/2} + |\mathbb{E}[f_n(\theta) - \tilde{f}_n(\theta)]|$.

We already have (8.21) for the second term. Now, as $\Theta_{m,p}$ goes to 0 if we take $m \to \infty$, from (8.22), we can say that both first and last terms in the right-hand-side of the above inequality will go to 0 and that completes the proof. □

**Proof of Theorem 5.2.** This theorem is a particular case of the general quadratic forms of stationary random fields and we will make use of Lemma 8.3 to prove the theorem. Note that $nf_n(\theta)$ is of the form $S_n$ in the above-mentioned lemma with $b_{n,j} = K(j/B_n)$. Using the given conditions, one can now show that $\sum_j b_{n,j}^2 \sim B_n \kappa$ and $\sum_{i,j} b_{n,i-j}^2 \sim n^d B_n \kappa$ where $\kappa = \int_{-\infty}^{\infty} K^2(x)dx < \infty$. That means the conditions of the above lemma are satisfied and hence, we can say that

$$\sqrt{\frac{n}{B_n}} \cdot \frac{f_n(\theta) - \mathbb{E}(f_n(\theta))}{f_J(\theta)} \Rightarrow N(0, \kappa). \tag{8.23}$$

Then, in view of the fact that $\mathbb{E}(f_n(\theta)) = f_J(\theta) + o(1)$, theorem 5.1 and a simple application of Slutsky's theorem completes our proof. □

**Proof of Theorem 6.1.** At first, we will define a new covariance matrix using the actual mean (unknown) of the process. Suppose, each $X_i$ has mean $\mu$ and let us define a new banded covariance matrix estimate

$$\hat{\Sigma}^0_{n,B_n} = \left(\hat{\gamma}^0_{L_i-L_j} K\left(\frac{L_i-L_j}{B_n}\right)\right)_{1\le i,j\le n},$$

where $\hat{\gamma}^0_k = \frac{1}{m_k}\sum_{i,j:L_i-L_j=k}(X_{L_i}-\mu)(X_{L_j}-\mu)$.

This proof will mainly use the Gershgorin circle theorem, which states that every eigenvalue of a complex $n\times n$ matrix $A$ lies within at least one of the Gershgorin discs $D_i(a_{ii},\sum_{j\ne i}|a_{ij}|)$, where $a_{ij}$'s are the elements of the matrix $A$. Thus, if $\lambda_i$, for $i=1,\dots,n$ denote eigenvalues of $A$, then the spectral radius satisfies the following:

$$\rho(A) = \max_i |\lambda_i| \le \max_i\left(|a_{ii}| + \sum_{j\ne i}|a_{ij}|\right) = \max_i \sum_{j=1}^n |a_{ij}|\,.$$

Let us now consider the spectral radius of the matrix $\hat{\Sigma}^0_{n,B_n} - \mathbb{E}(\hat{\Sigma}^0_{n,B_n})$. The $(i,j)$-th element of this matrix is $K((L_i-L_j)/B_n)[\hat{\gamma}^0_{L_i-L_j} - \mathbb{E}(\hat{\gamma}^0_{L_i-L_j})]$. Below, we will use $J_n$ and $J_{n,B_n}$ to denote the following two sets:

$$\begin{aligned} J_n &= \{k: \text{ there exists } 1\le i,j\le n \text{ satisfying } L_i-L_j=k\},\\ J_{n,B_n} &= \{k: |k|\le B_n \text{ and there exists } 1\le i,j\le n \text{ satisfying } L_i-L_j=k\}.\end{aligned}$$

Also, let us denote $J'_{n,B_n} = J_n - J_{n,B_n}$. Then, using the above mentioned property of spectral radius, it can be written that

$$\begin{aligned} &\rho\big(\hat{\Sigma}^0_{n,B_n} - \mathbb{E}(\hat{\Sigma}^0_{n,B_n})\big)\\ \le\;& \max_i \sum_{j=1}^n \left|\hat{\gamma}^0_{L_i-L_j}K((L_i-L_j)/B_n) - \mathbb{E}[\hat{\gamma}^0_{L_i-L_j}K((L_i-L_j)/B_n)]\right|\\ \le\;& 2\sum_{k\in J_n}\left|K\left(\frac{k}{B_n}\right)\right| \left\|\hat{\gamma}^0_k - \mathbb{E}(\hat{\gamma}^0_k)\right\|\\ \le\;& 2\sum_{k\in J_{n,B_n}} \left\|\hat{\gamma}^0_k - \mathbb{E}(\hat{\gamma}^0_k)\right\|. \end{aligned} \tag{8.24}$$

Note that if we relabel $X_i-\mu$ as $Y_i$ then the above expression essentially deals with the autocovariance function and its estimate of a zero-mean stationary process. Thus, if each $X_i\in \mathbb{L}^p$ for some $p\in(2,4]$, then based on the results obtained by Wu and Pourahmadi (2009) and using the assumption $m_k \asymp n$, we can write that the above bound is of the order $O_P(B_n^d n^{2/p-1}\Theta^2_{0,p})$. If we further assume that the process is $p$-stable (see Definition 2.2) which tells us that the term $\Theta_{0,p}$ is finite, we can say that the bound is of the order $O_P(B_n^d n^{2/p-1})$.

Now, for the covariance matrix estimate $\hat{\Sigma}_{n,B_n}$, we can write $\rho\big(\hat{\Sigma}_{n,B_n} - \mathbb{E}(\hat{\Sigma}_{n,B_n})\big) \le \rho\big(\hat{\Sigma}_{n,B_n} - \hat{\Sigma}^0_{n,B_n}\big) + \rho\big(\hat{\Sigma}^0_{n,B_n} - \mathbb{E}(\hat{\Sigma}^0_{n,B_n})\big) + \rho\big(\mathbb{E}(\hat{\Sigma}_{n,B_n}) - \mathbb{E}(\hat{\Sigma}^0_{n,B_n})\big)$.

We have already got the limit of the second term. For the first and third term, we can follow similar procedures as before. In case of first term, using Gershgorin Circle Theorem, we will again get a bound of a sum of terms of the form $\hat{\gamma}_k - \hat{\gamma}_k^0$, over the set $J_{n,B_n}$. Now, since $\bar{X} - \mu$ is $O_P(n^{-1})$, we can say that this bound is of the order $O_P(B_n^d n^{-1})$. Note that this bound is less than what we got for the second term. We can get similar results for the third term too.

Clearly, the overall bound for the spectral radius of $\hat{\Sigma}_{n,B_n} - \mathbb{E}(\hat{\Sigma}_{n,B_n})$ is $O_P(B_n^d n^{2/p-1})$. □

### *8.2. Proofs of lemmas*

**Lemma 8.1** (This lemma is due to El Machkouri et al. (2013)). *Following the aforementioned notation, consider $\Gamma_n$ and let $(\alpha_i)_{i\in\Gamma_n}$ be a family of real numbers. Then, for any $p \geq 2$, we get*

$$\Big\|\sum_{i\in\Gamma_n} \alpha_i X_i\Big\|_p \leq \Big(2p \sum_{i\in\Gamma_n} \alpha_i^2\Big)^{1/2} \Delta_p. \tag{8.25}$$

**Lemma 8.2.** *Let $X_i \in L^p$ for $p > 1$, $\mathbb{E}(X_k) = 0$, $\alpha_1, \alpha_2, \ldots \in \mathbb{C}$, $p' = \min\{2, p\}$, $A_n = (\sum_{k=1}^n |\alpha_{L_k}|^{p'})^{1/p'}$ and $C_p = 18p^{3/2}(p-1)^{-1/2}$. Then, $\|\sum_{k=1}^n \alpha_{L_k} X_{L_k}\|_p \leq C_p A_n \Theta_{0,p}$, $\left\|\sum_{k=1}^n \alpha_{L_k} \tilde{X}_{L_k}\right\|_p \leq C_p A_n \Theta_{0,p}$ and $\left\|\sum_{k=1}^n \alpha_{L_k} (X_{L_k} - \tilde{X}_{L_k})\right\|_p \leq C_p A_n \Theta_{m+1,p}$.*

*Proof.* Let $\tau : \mathbb{Z} \to \mathbb{Z}^d$ be a bijection. For any $i \in \mathbb{Z}$, for any $j \in \mathbb{Z}^d$, define the projection operator $P_i X_j := \mathbb{E}(X_j \mid \mathcal{F}_i) - \mathbb{E}(X_j \mid \mathcal{F}_{i-1})$, where $\mathcal{F}_i = \sigma(\varepsilon_{\tau(l)}; l \leq i)$. Also, define $T^j \mathcal{F}_i = \sigma(\varepsilon_{\tau(l)-j}; l \leq i)$. Now, it is easy to note that $\|P_i X_j\| \leq \|X_{j-\tau(i)} - X^*_{j-\tau(i)}\|_p$ and hence, we get the inequality $\|P_i X_j\| \leq \delta_{j-\tau(i),p}$.

Then, noting that $X_i = \sum_{j\in\mathbb{Z}} P_j X_i$ for all $i \in \mathbb{Z}^d$, we can make use of Burkholder inequality and Cauchy-Schwarz inequality to get the first two results. The proof here is similar to Proposition 1 in El Machkouri et al. (2013). And then, the third result is a direct consequence of the first two. □

**Lemma 8.3.** *Let $\beta_j \in \mathbb{R}$ with $\beta_j = \beta_{-j}$; $\alpha_j = \beta_j e^{\imath j'\theta}$ where $\theta \in [-\pi, \pi]^d$. Also, for any $\theta$, define $\bar{\omega}(\theta) = 2$ if $\theta/\pi \in \mathbb{Z}^d$. Else, define it to be 1. Consider the quadratic form*

$$S_n = \sum_{1\leq j,k\leq n} \alpha_{L_k - L_j} X_{L_j} X_{L_k} \qquad \text{and} \qquad \sigma_n^2 = \bar{\omega}(\theta) \sum_{1\leq j,k\leq n} \beta^2_{L_j - L_k}, \tag{8.26}$$

*where we assume that $\mathbb{E}(X_0) = 0, X_0 \in L^4, \Theta_{0,4} < \infty$. In addition, let $\zeta_n^2 = \sum_{1\leq t\leq n} \beta^2_{L_t}$, and let us assume that $\max_{1\leq t\leq n} \beta^2_{L_t} = o(\zeta_n^2), n^d \zeta_n^2 = O(\sigma_n^2)$. We further consider that*

$$\sum_{k=1}^n \sum_{t=1}^{k-1} \Big| \sum_{j=1+k}^n \alpha_{L_k - L_j} \alpha_{n, L_t - L_j} \Big|^2 = o(\sigma_n^4), \ \sum_{k=1}^n \big|\beta_{L_k} - \beta_{L_k - 1}\big|^2 = o(\zeta_n^2). \tag{8.27}$$

*Now, following Section 3.2, suppose $J$ denotes the set $\{k \in \mathbb{Z}^d : \exists\, i, j$ with $L_i - L_j = k\}$ and $m_k$ is the cardinality of the set $\{(i,j) : L_i - L_j = k\}$. If $f_J(\theta)$ denotes the location adjusted spectral density, then*

$$\frac{S_n - \mathbb{E}(S_n)}{\sigma_n f_J(\theta)} \xrightarrow[n\to\infty]{L} N\big(0,1\big). \tag{8.28}$$

*Proof.* We will make use of $m$-dependence approximation, defined in 8.18, to prove this result. Once again, for simplicity, we prove the result for $d = 2$, but the idea can be easily used for higher dimensions in a similar fashion. To begin with, note that $S_n$ is a special case of $U_n = \sum_{s,t\in\mathbb{Z}_n^d} a_{s,t} X_s X_t$, where $\mathbb{Z}_n^d$ is a regular grid and $a_{s,t}$ are appropriate constants. Further, we have, $\big\|\sum_s X_s^2 - n^d\gamma_0\big\| = O(n^{d/2})$. Then, defining $Z_s = \sum_{t\neq s} a_{s,t} X_t$ and $\tilde{Z}_s = \sum_{t\neq s} a_{s,t}\tilde{X}_t$, we write $T_n = \sum_s X_s Z_s$, $\tilde{T}_n = \sum_s \tilde{X}_s \tilde{Z}_s$ and $T_n^* = \sum_s X_s \tilde{Z}_s$.

Using the previous lemma and with similar arguments as in Proposition 1 of Liu and Wu (2010), we can show that

$$\frac{\big\|\big(T_n - \mathbb{E}(T_n)\big) - \big(T_n^* - \mathbb{E}(T_n^*)\big)\big\|_p}{n^{d/2}\zeta_n} \le C_p d_m, \tag{8.29}$$

where $d_m \to 0$ as $m \to \infty$. We can get a similar inequality for $\big\|(\tilde{T}_n - \mathbb{E}(\tilde{T}_n)) - (T_n^* - \mathbb{E}(T_n^*)\big\|_p^2$ and then, using both inequalities, we can write that $\|(T_n - \mathbb{E}(T_n)) - (\tilde{T}_n - \mathbb{E}(\tilde{T}_n))\| = o(\sigma_n)$. Thus, in order to get the required result, it is enough to find the asymptotic distribution of $(\tilde{T}_n - \mathbb{E}(\tilde{T}_n))/\sigma_n$.

We write $\tilde{T}_n = \sum_{s,t:|s-t|<2m} a_{s,t}\tilde{X}_s\tilde{X}_t + \sum_{s,t:|s-t|\ge 2m} a_{s,t}\tilde{X}_s\tilde{X}_t$. It is easy to note that the first term is $O(n^{d/2}\max|a_{s,t}|) = o(\sigma_n)$. For the second term, writing it using the coordinates of the indexes and using the previously defined sigma fields, we will make use of the martingale central limit theorem (Hall and Heyde (2014)) and that will give us the required result directly. The arguments here will be similar to the one dimensional case, as was done in Liu and Wu (2010). ☐

**Lemma 8.4.** *Suppose $I_n(\theta)$ denotes the periodogram corresponding to the data $(X_{L_i})_{i=1(1)n}$ where each $L_i$ is an element from $\mathbb{R}^d$. Then, for a square integrable function $f(\theta)$ defined on $D \subset \mathbb{R}^d$, variance of the quantity $\int_D I_n(\theta) f(\theta)\, d\theta$ goes to 0, as $n$ goes to $\infty$.*

*Proof.* Note that $nI_n(\theta)$ is a quadratic form in $X_{L_j}$'s and one can get that $\int_D nI_n(\theta)f(\theta)\,d\theta = \sum_{j,k}\alpha_{L_j-L_k}X_{L_j}X_{L_k}$ where $\alpha_k$ is same as the Fourier coefficients corresponding to the function $f(\theta)$. If we denote it by $nT_n$ and define $n\tilde{T}_n$ in a similar way, but with $\tilde{X}_{L_j}$'s, then using similar arguments as in the previous lemma, one can write that

$$\Big\|(S_n - \mathbb{E}(S_n)) - (\tilde{S}_n - \mathbb{E}(\tilde{S}_n))\Big\| = O_p(A_n/\sqrt{n}),$$

where $A_n = (\sum_{k=1}^n |\alpha_{L_k}|^2)^{1/2}$. Since $\alpha_k$ denotes the Fourier coefficient corresponding to $f(\theta)$ and since $f$ is square integrable, we can say that $A_n = O_p(1)$.

Now, $n\tilde{S}_n$ is obtained using the $m$-dependent approximations and so, for a fixed $m$, the variance of $\tilde{S}_n$ goes to 0. Combining the above, it is straightforward to show that the variance of $\int_D I_n(\theta)f(\theta)\, d\theta$ (which is same as $S_n$) goes to 0 as $n \to \infty$. □

## Acknowledgement

The authors thank two anonymous referees and Editor Domenico Marinucci for their useful suggestions during the review process. The work of the second author was supported by the National Science Foundation DMS-1612984.

## References

T. T. Azomahou. Memory properties and aggregation of spatial autoregressive models. *Journal of Statistical Planning and Inference*, 139(8):2581–2597, 2009. MR2523650

S. Bandyopadhyay and S. N. Lahiri. Asymptotic properties of discrete Fourier transforms for spatial data. *Sankhyā*, 71(2, Ser. A):221–259, 2009. ISSN 0972-7671. MR2639292

S. Bandyopadhyay, S. N. Lahiri, and D. J. Nordman. A frequency domain empirical likelihood method for irregularly spaced spatial data. *The Annals of Statistics*, 43(2):519–545, 2015. ISSN 0090-5364. MR3316189

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236, 1974. MR0373208

W. Clinger and J. W. Van Ness. On unequally spaced time points in time series. *The Annals of Statistics*, 4(4):736–745, 1976. ISSN 0090-5364. MR0478514

N. Cressie. Spatial prediction and ordinary kriging. *Mathematical geology*, 20 (4):405–421, 1988. MR0950457

N. Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015. MR3559472

R. Dahlhaus and H. Künsch. Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, 74(4):877–882, 1987. ISSN 0006-3444. MR0919857

M. El Machkouri, D. Volný, and W. B. Wu. A central limit theorem for stationary random fields. *Stochastic Processes and their Applications*, 123(1):1–14, 2013. ISSN 0304-4149. MR2988107

M. Fuentes. Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, 102(477):321–331, 2007. ISSN 0162-1459. MR2345545

X. Guyon. Parameter estimation for a stationary process on a $d$-dimensional lattice. *Biometrika*, 69(1):95–105, 1982. ISSN 0006-3444. MR0655674

P Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic press, 2014. MR0624435

P. Hall and P. Patil. Properties of nonparametric estimators of autocovariance for stationary random fields. *Probability Theory and Related Fields*, 99(3): 399–424, 1994. ISSN 0178-8051. MR1283119

L. Heinrich. Asymptotic behaviour of an empirical nearest-neighbour distance function for stationary Poisson cluster processes. *Mathematische Nachrichten*, 136:131–148, 1988. ISSN 0025-584X. MR0952468

A. A. Ivanov and N. Leonenko. *Statistical analysis of random fields*, volume 28. Springer Science & Business Media, 2012. MR1009786

R. H. Jones. Spectral analysis with regularly missed observations. *Annals of Mathematical Statistics*, 33:455–461, 1962. ISSN 0003-4851. MR0139250

R. L. Kashyap. Characterization and estimation of two-dimensional ARMA models. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, 30(5):736–745, 1984. ISSN 0018-9448. MR0781270

F. Lavancier. Aggregation of isotropic autoregressive fields. *Journal of Statistical Planning and Inference*, 141(12):3862–3866, 2011. MR2823655

B. Li, M. G. Genton, and M. Sherman. On the asymptotic joint distribution of sample space-time covariance estimators. *Bernoulli*, 14(1):228–248, 2008. ISSN 1350-7265. MR2401661

Z. Lin and W. Liu. On maxima of periodograms of stationary processes. *The Annals of Statistics*, 37(5B):2676–2695, 2009. ISSN 0090-5364. MR2541443

W. Liu and W. B. Wu. Asymptotics of spectral density estimates. *Econometric Theory*, 26(4):1218–1245, 2010. ISSN 0266-4666. MR2660298

Y. Matsuda and Y. Yajima. Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 71 (1):191–217, 2009. ISSN 1369-7412. MR2655530

H. R. Neave. Spectral analysis of a stationary time series using initially scarce data. *Biometrika*, 57:111–122, 1970a. ISSN 0006-3444. MR0275615

H. R. Neave. An improved formula for the asymptotic variance of spectrum estimates. *Annals of Mathematical Statistics*, 41:70–77, 1970b. ISSN 0003-4851. MR0251876

E. Parzen. On spectral analysis with missing observations and amplitude modulation. *Sankhyā Ser. A*, 25:383–392, 1963. ISSN 0581-572X. MR0172435

M. Peligrad and W. B. Wu. Central limit theorem for Fourier transforms of stationary processes. *The Annals of Probability*, 38(5):2009–2022, 2010. ISSN 0091-1798. MR2722793

T. Pukkila. The bias in periodogram ordinates and the estimation of ARMA models in the frequency domain. *The Australian Journal of Statistics*, 21(2): 121–128, 1979. ISSN 0004-9581. MR0547387

J. M. Vidal-Sanz. Automatic spectral density estimation for random fields on a lattice via bootstrap. *TEST*, 18(1):96–114, 2009. ISSN 1133-0686. MR2495966

P. Whittle. On stationary processes in the plane. *Biometrika*, 41:434–449, 1954. ISSN 0006-3444. MR0067450

W. B. Wu. Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (40):14150–14154 (electronic), 2005. ISSN 1091-6490. MR2172215

W. B. Wu and M. Pourahmadi. Banding sample autocovariance matrices of stationary processes. *Statistica Sinica*, 19(4):1755–1768, 2009. ISSN 1017-0405. MR2589209