STATISTICAL INFERENCE FOR SPATIAL STATISTICS DEFINED IN THE FOURIER DOMAIN

BY SUHASINI SUBBA RAO[1]

Texas A&M University

A class of Fourier based statistics for irregular spaced spatial data is introduced. Examples include the Whittle likelihood, a parametric estimator of the covariance function based on the L_2-contrast function and a simple nonparametric estimator of the spatial autocovariance which is a nonnegative function. The Fourier based statistic is a quadratic form of a discrete Fourier-type transform of the spatial data. Evaluation of the statistic is computationally tractable, requiring $O(nb)$ operations, where b are the number of Fourier frequencies used in the definition of the statistic and n is the sample size. The asymptotic sampling properties of the statistic are derived using both increasing domain and fixed-domain spatial asymptotics. These results are used to construct a statistic which is asymptotically pivotal.

1. Introduction. In recent years, irregular spaced spatial data has become ubiquitous in several disciplines as varied as the geosciences to econometrics. The analysis of such data poses several challenges which do not arise in data which is sampled on a regular lattice. A major obstacle is the computational costs when dealing with large irregular sampled data sets. If spatial data are sampled on a regular lattice, then algorithms such as the fast Fourier transform can be employed to reduce the computational burden (see, e.g., [9]). Unfortunately, such algorithms have little benefit if the spatial data are irregularly sampled. To address this issue, within the spatial domain, several authors, including [10, 35, 38], have proposed estimation methods which are designed to reduce the computational burden.

In contrast to the above references, [17] and [27] argue that working within the frequency domain often simplifies the computational burden. Both authors focus on parametric estimation using a Whittle-type likelihood. Fuentes [17] assumes that the irregular spaced data can be embedded on a grid and the missing mechanism is deterministic and "locally smooth." A possible drawback of this construction is that the local smooth assumption will not hold if the locations are extremely irregular. Therefore, [27] proposes a Whittle likelihood approach to parameter estimation which takes into account the irregular nature of the locations. The focus of most Fourier domain estimators have been on the Whittle likelihood (the exception

being the recent paper by [2], which we discuss later). In this paper, we argue that
several estimators, both parametric and nonparametric, can be defined within the
Fourier domain. For example, within the Fourier domain, we propose a nonpara-
metric, nonnegative definite estimator of the spatial covariance. Nonparametric
estimators of the spatial autocovariance are often defined using kernel smoothing
methods (see [20]) or the empirical variogram (see [11]). However, these "raw"
covariance estimators may not be nonnegative functions, and a second step is re-
quired, which involves taking the Fourier transform of a finite discretisation of the
sample autocovariance, setting negative values to zero and inverting back, to en-
sure that the resulting estimator is a nonnegative function. In contrast, by defining
the covariance estimator within the Fourier domain the estimator is guaranteed to
be a nonnegative definite function. The purpose of this paper is twofold. The first
is to demonstrate that several parameters can be estimated within the Fourier do-
main. The second is to obtain a comprehensive understanding of quadratic forms
of irregular sampled spatial processes.

In order to define estimators within the Fourier domain, we adopt the approach
pioneered by [27] and [1] who assume that the irregular locations are independent,
identically distributed random variables (thus allowing the data to be extremely
irregular) and define the irregular sampled discrete Fourier transform (DFT) as

$$(1.1) \qquad J_n(\boldsymbol{\omega}) = \frac{\lambda^{d/2}}{n} \sum_{j=1}^{n} Z(\boldsymbol{s}_j) \exp(i\boldsymbol{s}_j' \boldsymbol{\omega}),$$

where $\boldsymbol{s}_j \in [-\lambda/2, \lambda/2]^d$ denotes the spatial locations observed in the space
$[-\lambda/2, \lambda/2]^d$ and $\{Z(\boldsymbol{s}_j)\}$ denotes the spatial random field at these locations. It
is worth mentioning a similar transformation on irregular sampled data goes back
to [26], who defines the discrete Fourier transform of Poisson sampled continuous
time series. Using this definition, [27] define the Whittle likelihood by taking the
weighted integral of the periodogram, $|J_n(\boldsymbol{\omega})|^2$. Of course, in practice the weighted
integral needs to be approximated by a Riemann sum. Indeed in Remark 2, [27]
suggest using the frequency grid $\{\boldsymbol{\omega_k} = 2\pi \boldsymbol{k}/\lambda; \boldsymbol{k} \in \mathbb{Z}^d\}$ when constructing the
Whittle likelihood. No justification is given for this discretisation. However, their
observation has insight. We prove that this transformation is "optimal" for most
estimators defined within the Fourier domain.

Motivated by the integrated Whittle likelihood, our aim is to consider estimators
with the form $\int g_\theta(\boldsymbol{\omega})|J_n(\boldsymbol{\omega})|^2 \, d\boldsymbol{\omega}$. Such quantities have been widely studied in
time series, dating as far back as [29], but has received very little attention in the
spatial literature. In practice, this integral cannot be evaluated, and needs to be
approximated by a Riemann sum:

$$(1.2) \qquad Q_{a,\Omega,\lambda}(g_\theta; 0) = \frac{1}{\Omega^d} \sum_{k_1,\ldots,k_d=-a}^{a} g_\theta(\boldsymbol{\omega}_{\Omega,\boldsymbol{k}})|J_n(\boldsymbol{\omega}_{\Omega,\boldsymbol{k}})|^2,$$

where $\{\boldsymbol{\omega}_{\Omega,\boldsymbol{k}} = 2\pi\boldsymbol{k}/\Omega, \boldsymbol{k} = (k_1, \ldots, k_d), -a \leq k_i \leq a\}$ is the frequency grid over which the sum is evaluated. In terms of computation, evaluation of $\{J_n(\boldsymbol{\omega}_{\Omega,\boldsymbol{k}});$ $\boldsymbol{k} = (k_1, \ldots, k_d), k_j = -a, \ldots, a\}$ requires $O(a^d n)$ operations. However, once $\{J_n(\boldsymbol{\omega}_{\Omega,\boldsymbol{k}})\}$ has been evaluated the evaluation of $Q_{a,\Omega,\lambda}(g_\theta; 0)$ only requires $O(a^d)$ operations.

As far as we are aware, there exists no results on the sampling properties of the general quadratic form defined in (1.2). To derive the asymptotic sampling properties of $Q_{a,\Omega,\lambda}(g_\theta; 0)$, we will work under two asymptotic frameworks that are commonly used in spatial statistics. Our main focus will be the increasing domain framework, introduced in [21] (see also [20] and used in, e.g., [1, 2, 24, 27] and [3]). This is where the number of observed locations $n \to \infty$ as the size of the spatial domain $\lambda \to \infty$ (we usually assume $\lambda^d/n \to 0$). We also analyze the sampling properties of $Q_{a,\Omega,\lambda}(g_\theta; 0)$ within the fixed-domain framework (where λ is kept fixed but the number of locations, n grows) considered in [33, 34, 40] and [41]. The sampling properties of $Q_{a,\Omega,\lambda}(g_\theta; 0)$ differ according to the framework used.

We show in Sections 3 and 4 that $Q_{a,\Omega,\lambda}(g_\theta; 0)$ is a consistent estimator of the functional $I(g_\theta; \frac{a}{\Omega})$ as $\lambda \to \infty$ and $\Omega \to \infty$, where

$$(1.3) \qquad I\left(g_\theta; \frac{a}{\Omega}\right) = \frac{1}{(2\pi)^d} \int_{[-2\pi a/\Omega, 2\pi a/\Omega]^d} g_\theta(\boldsymbol{\omega}) f(\boldsymbol{\omega}) \, d\boldsymbol{\omega}.$$

However, the choice of frequency grid $\boldsymbol{\omega}_{\Omega,\boldsymbol{k}}$ plays a vital role in the rate of convergence. In particular, we show that

$$\mathrm{E}[Q_{a,\Omega,\lambda}(g_\theta; 0)] = I\left(g_\theta; \frac{a}{\Omega}\right) + O\left(\frac{\log \lambda}{\lambda} + \frac{1}{\Omega} + \frac{1}{n}\right)$$

and

$$(1.4) \qquad \mathrm{var}[Q_{a,\Omega,\lambda}(g_\theta; 0)] = \begin{cases} O(\lambda^{-d}), & \Omega > \lambda, \\ O(\Omega^{-d}), & \Omega \leq \lambda. \end{cases}$$

Therefore, under suitable conditions on g_θ, $Q_{a,\Omega,\lambda}(g_\theta; 0) \xrightarrow{\mathcal{P}} I(g_\theta; \infty)$ if $a/\Omega \to \infty$ as $a \to \infty$ $n \to \infty$, $\lambda \to \infty$ and $\Omega \to \infty$.

To understand the influence the user chosen frequency grid has on the sampling properties, we show that the asymptotic limit of $\lambda^d \mathrm{var}[Q_{a,\Omega,\lambda}(g_\theta; 0)]$ will always be the same for all $\Omega \geq \lambda$ as $\lambda \to \infty$. On the other hand, using a frequency grid which is coarser than $\{2\pi\boldsymbol{k}/\lambda; \boldsymbol{k} \in \mathbb{Z}^d\}$ leads to an estimator with a larger bias and variance. Thus, balancing efficiency with computational burden, in general, $\{J_n(\boldsymbol{\omega}_{\lambda,\boldsymbol{k}})\}_{\boldsymbol{k} \in \mathbb{Z}^d}$ is the optimal transformation of the spatial data into the frequency domain. As mentioned above, [2] also use the Fourier domain for spatial inference; however, their objectives are very different to those in this paper. Bandyopadhyay et al. [2] show that the transformations $\{J_n(\boldsymbol{\omega}_{\Omega,\boldsymbol{k}})\}_{\boldsymbol{k}}$ are asymptotically independent if $\Omega/\lambda \to 0$ as $\lambda \to \infty$ (this corresponds to a very coarse frequency grid).

Based on this property, they use $Q_{a,\Omega,\lambda}(g;0)$, where Ω is such that $\Omega/\lambda \to 0$ as $\lambda \to \infty$ and $\Omega \to \infty$, to construct the empirical likelihood. The justification for their construction is that the distribution of the resulting empirical likelihood is asymptotically pivotal as $\lambda \to \infty$. The sampling properties of $Q_{a,\Omega,\lambda}(g;0)$ are not derived in [2]. However, it is clear from (1.4), that using a frequency grid where $\Omega \ll \lambda$ leads to an estimator that is not optimal in the mean squared sense.

Since $Q_{a,\Omega,\lambda}(g;0)$ is optimal when using the frequency grid $\Omega = \lambda$, in Section 4 we focus on deriving the sampling properties of $Q_{a,\lambda,\lambda}(g_\theta;0)$. We consider the slightly more general statistic

$$(1.5) \quad Q_{a,\lambda,\lambda}(g_\theta; \boldsymbol{r}) = \frac{1}{\lambda^d} \sum_{k_1,\ldots,k_d=-a}^{a} g_\theta(\boldsymbol{\omega}_{\lambda,\boldsymbol{k}}) J_n(\boldsymbol{\omega}_{\lambda,\boldsymbol{k}}) \overline{J_n(\boldsymbol{\omega}_{\lambda,\boldsymbol{k+r}})}, \qquad \boldsymbol{r} \in \mathbb{Z}^d,$$

and show asymptotic normality of $Q_{a,\lambda,\lambda}(g_\theta; \boldsymbol{r})$ when the random field is stationary and Gaussian and obtain the second-order properties of $Q_{a,\lambda,\lambda}(g_\theta; \boldsymbol{r})$ when the random field is stationary (but not necessarily Gaussian). The sampling properties of $Q_{a,\lambda,\lambda}(g_\theta;0)$ when the domain is kept fixed are considered in Section 4.4. The variance of $Q_{a,\lambda,\lambda}(g_\theta;0)$ is usually difficult to directly estimate. However, in Section 5 we show that if the locations are independent, uniformly distributed random variables, then $\{Q_{a,\lambda,\lambda}(g_\theta; \boldsymbol{r})\}$ forms a "near uncorrelated" sequence whose variance is asymptotically equivalent to $Q_{a,\lambda,\lambda}(g_\theta;0)$. More precisely, if $Q_{a,\lambda,\lambda}(g_\theta;0)$ is real we define the the Studentized statistic

$$(1.6) \qquad T_{\mathcal{S}} = \frac{Q_{a,\lambda,\lambda}(g_\theta;0) - I(g_\theta; \frac{a}{\lambda})}{\sqrt{\frac{1}{|\mathcal{S}|} \sum_{\boldsymbol{r} \in \mathcal{S}} |Q_{a,\lambda,\lambda}(g_\theta; \boldsymbol{r})|^2}},$$

for some fixed set $\mathcal{S} \subset \mathbb{Z}^d/\{0\}$. We show that $T_{\mathcal{S}} \xrightarrow{\mathcal{D}} t_{2|\mathcal{S}|}$ as $\lambda \to \infty$, where $t_{2|\mathcal{S}|}$ denotes a t-distribution with $2|\mathcal{S}|$ degrees of freedom and $|\mathcal{S}|$ denotes the cardinality of $\mathcal{S}$.

We now summarize the paper. In Section 2, we state the assumptions required in this paper and the sampling properties of the Fourier transform $\{J_n(\boldsymbol{\omega}_{\Omega,\boldsymbol{k}})\}$. In Section 2.3, we use these properties to motivate examples of estimators which have the form $Q_{a,\Omega,\lambda}(g_\theta;0)$. In Section 3, we summarize the sampling properties of $\{Q_{a,\Omega,\lambda}(g_\theta;0)\}$. In Section 4, we focus on $Q_{a,\lambda,\lambda}(g_\theta; \boldsymbol{r})$ and these results are used to study the sampling properties of $T_{\mathcal{S}}$ in Section 5. $Q_{a,\Omega,\lambda}(g_\theta;0)$ is a quadratic form of an irregular sampled spatial process, and as far as we are aware, there exist very few results on the moment and sampling properties of such quadratic forms. The purpose of the supplementary material, [36], is to take a few steps in this direction. Many of these results build on the work of [23], [6], [14] and [30] and may be of independent interest. A simulation study to illustrate the performance of the nonparametric nonnegative definite estimator of the spatial covariance is given in [36], Appendix J.

2. Assumptions and examples.

2.1. *Assumptions and notation.* In this section we state the required assumptions and notation. This section can be skipped on the first reading.

We observe the spatial random field $\{Z(s); s \in \mathbb{R}^d\}$ at the locations $\{s_j\}_{j=1}^n$ where $s_j \in [-\lambda/2, \lambda/2]^d$. Throughout this paper, we will use the following assumptions on the spatial random field.

ASSUMPTION 2.1 (Spatial random field). (i) $\{Z(s); s \in \mathbb{R}^d\}$ is a second-order stationary random field with mean zero and covariance function: $c(s_1 - s_2) = \operatorname{cov}(Z(s_1), Z(s_2)|s_1, s_2)$. We define the spectral density function as $f(\omega) = \int_{\mathbb{R}^d} c(s) \exp(-is'\omega)\,ds$ [and $c(s) = (2\pi)^{-d} \int_{\mathbb{R}^d} f(\omega) \exp(is'\omega)\,d\omega$].
(ii) $\{Z(s); s \in \mathbb{R}^d\}$ is a stationary Gaussian random field.

We require the following definitions. For some finite $0 < C < \infty$ and $\delta > 0$, let

$$(2.1) \qquad \beta_\delta(s) = \begin{cases} C, & |s| \in [-1, 1], \\ C|s|^{-\delta}, & |s| > 1. \end{cases}$$

Let $\beta_\delta(s) = \prod_{j=1}^d \beta_\delta(s_j)$. For sequences, we define $\xi_\eta(j) = C[I(j = 0) + I(j \neq 0)|j|^{-\eta}]$ [for some finite constant C and $I(\cdot)$ denotes the indicator function]. To minimise notation, we will often use $\sum_{k=-a}^a$ to denote the multiple sum $\sum_{k_1=-a}^a \cdots \sum_{k_d=-a}^a$. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the ℓ_1-norm and ℓ_2-norm of a vector, respectively. Let $\Re X$ and $\Im X$ denote the real and imaginary parts of X. We make heavy use of the sinc function which is defined as

$$(2.2) \qquad \operatorname{sinc}(\omega) = \frac{\sin(\omega)}{\omega} \quad \text{and} \quad \operatorname{Sinc}(\boldsymbol{\omega}) = \prod_{j=1}^d \operatorname{sinc}(\omega_j).$$

Define the triangle kernel, $T : \mathbb{R} \to \mathbb{R}$ where $T(u) = 1 - |u|$ for $u \in [-1, 1]$ and zero elsewhere and the d-dimensional triangle kernel $T(\boldsymbol{u}) = \prod_{j=1}^d T(u_j)$. We use the notation $\{\boldsymbol{\omega}_{\Omega,k} = 2\pi \boldsymbol{k}/\Omega; \boldsymbol{k} \in \mathbb{Z}^d\}$ for a general frequency grid. If $\Omega > \lambda$, we say the frequency grid is "fine." Conversely, if $\Omega < \lambda$, we say the frequency grid is "coarse." Further, as mentioned in the Introduction, using $\lambda = \Omega$ is optimal, therefore, to reduce notation we let $\{\boldsymbol{\omega}_k = 2\pi \boldsymbol{k}/\lambda; \boldsymbol{k} \in \mathbb{Z}^d\}$.

We adopt the assumptions of [21, 27] and [1] and assume that $\{s_j\}$ are i.i.d. random variables with density $\frac{1}{\lambda^d} h(\frac{\cdot}{\lambda})$, where $h : [-1/2, -1/2]^d \to \mathbb{R}$.

ASSUMPTION 2.2 (Nonuniform sampling). The locations $\{s_j\}$ are independent distributed random variables on $[-\lambda/2, \lambda/2]^d$, where the density of $\{s_j\}$ is $\frac{1}{\lambda^d} h(\frac{\cdot}{\lambda})$, and $h(\cdot)$ admits the Fourier representation

$$h(\boldsymbol{u}) = \sum_{\boldsymbol{j} \in \mathbb{Z}^d} \gamma_{\boldsymbol{j}} \exp(i2\pi \boldsymbol{j}'\boldsymbol{u}),$$

where $\sum_{\boldsymbol{j} \in \mathbb{Z}^d} |\gamma_{\boldsymbol{j}}| < \infty$ such that $|\gamma_{\boldsymbol{j}}| \leq C \prod_{i=1}^{d} \xi_{1+\delta}(j_i)$ (for some $\delta > 0$). This assumption is satisfied if the second derivative of h is bounded on the d-dimensional torus $[-1/2, 1/2]^d$.

REMARK 2.1.   If h is such that $\sup_{\boldsymbol{s} \in [-1/2, 1/2]^d} |\frac{\partial^{m_1 + \cdots , m_d} h(s_1, \ldots, s_d)}{\partial s_1^{m_1} \cdots \partial s_d^{m_d}}| < \infty$ ($0 \leq m_i \leq 2$) but h is not continuous on the d-dimensional torus $[-1/2, 1/2]^d$, then $|\gamma_{\boldsymbol{j}}| \leq C \prod_{i=1}^{d} \xi_1(j_i)$ and the above condition will not be satisfied. However, this assumption can be induced by tapering the observations such that $Z(\boldsymbol{s}_j)$ is replaced with $\widetilde{Z}(\boldsymbol{s}_j)$, where $\widetilde{Z}(\boldsymbol{s}_j) = t(\boldsymbol{s}_j) Z(\boldsymbol{s}_j)$, $t(\boldsymbol{s}) = \prod_{i=1}^{d} t(s_i)$ and t is a weight function which has a bounded second derivative, $t(-1/2) = t(1/2) = 0$ and $t'(1/2) = t'(-1/2) = 0$. By using $\widetilde{Z}(\boldsymbol{s}_j)$ instead of $Z(\boldsymbol{s}_j)$, in all the derivations below we replace the density $h(\boldsymbol{s})$ with $t(\boldsymbol{s})h(\boldsymbol{s})$. This means the results now rely on the Fourier coefficients of $t(\boldsymbol{s})h(\boldsymbol{s})$, which decay at the rate $|\int_{[-1/2, 1/2]^d} t(\boldsymbol{s})h(\boldsymbol{s}) \exp(i2\pi \boldsymbol{j}'\boldsymbol{s}) \, d\boldsymbol{s}| \leq C \prod_{i=1}^{d} \xi_2(j_i)$, and thus the above condition is satisfied. Note that [27], Definition 2, uses a similar data-tapering scheme to induce a similar condition.

The case that the locations follow a uniform distribution is an example of a distribution which satisfies Assumption 2.2. It gives rise to several elegant simplifications. Thus, we state the uniform case as a separate assumption.

ASSUMPTION 2.3 (Uniform sampling).   The locations $\{\boldsymbol{s}_j\}$ are independent uniformly distributed random variables on $[-\lambda/2, \lambda/2]^d$.

Many of the results in this paper use that the locations follow a random design. This helps in understanding the sampling properties of these complex estimators. However, it can "mask" the approximation errors when replacing sums by integrals and the role that the sample size n plays in these approximations. To get some idea of these approximations when the domain λ is kept fixed but $n \to \infty$, we will, on occasion, treat the locations as deterministic and make the following assumption.

ASSUMPTION 2.4 (Near lattice locations for $d = 1$).   Let $\{s_{n,j}; j = 1, \ldots, n\}$ denote the locations. The number of locations $n \to \infty$ in such a way that

$$\sum_{j=1}^{n-1} \left| \frac{\lambda}{n} - (s_{n,(j+1)} - s_{n,(j)}) \right| = O\left(\frac{\lambda}{n}\right),$$

$$\sum_{j=1}^{n-1} (s_{n,(j+1)} - s_{n,(j)})^2 = O\left(\frac{\lambda}{n}\right),$$

where $\{s_{n,(j)}\}_j$ denotes the order statistics corresponding $\{s_{n,j}\}_j$.

The integrated periodogram $Q_{a,\Omega,\lambda}(g;0)$ resembles the integrated periodogram estimator commonly used in time series (see, e.g., [8, 13, 15, 22, 39] and [28]). However, there are some fundamental differences, between time series estimators and $Q_{a,\Omega,\lambda}(g;0)$ which makes the analysis very different. Unlike regularly spaced or near regularly spaced data, "truly" irregular sampling means that the DFT can estimate high frequencies, without the curse of aliasing (a phenomena which was noticed as early as [32] and [4]). In this case, if the function g_θ, in the definition of $Q_{a,\Omega,\lambda}(g_\theta;0)$ is bounded, there is no need for the frequency grid to be bounded, and a can be magnitudes larger than λ. Below we state assumptions on the function g_θ and the frequency grid.

ASSUMPTION 2.5 [Assumptions on $g_\theta(\cdot)$ and the size of frequency grid]. Suppose

$$Q_{a,\Omega,\lambda}(g_\theta;r) = \frac{1}{\Omega^d} \sum_{k=-a}^{a} g_\theta(\boldsymbol{\omega}_{\Omega,k}) J_n(\boldsymbol{\omega}_{\Omega,k}) \overline{J_n(\boldsymbol{\omega}_{\Omega,k+r})}.$$

(i) If g_θ is not a bounded function over $\mathbb{R}^d$ but $\sup_{\boldsymbol{\omega}\in[-C,C]^d} |g_\theta(\boldsymbol{\omega})| < \infty$, then we must restrict the frequency grid $\{\boldsymbol{\omega}_{\Omega,k}; -a \leq k_1, \ldots, k_d \leq a\}$ to lie in $[-C, C]^d$ (thus $a = C\Omega$). Further, we assume for all $1 \leq j \leq d$, $\sup_{\boldsymbol{\omega}\in[-C,C]^d} |\frac{\partial g_\theta(\boldsymbol{\omega})}{\partial \omega_j}| < \infty$.

(ii) If $\sup_{\boldsymbol{\omega}\in\mathbb{R}^d} |g_\theta(\boldsymbol{\omega})| < \infty$, then the frequency grid can be unbounded (in the sense that $a/\Omega \to \infty$ as a and $\Omega \to \infty$). Further, we assume for all $1 \leq j \leq d$, $\sup_{\boldsymbol{\omega}\in\mathbb{R}^d} |\frac{\partial g_\theta(\boldsymbol{\omega})}{\partial \omega_j}| < \infty$.

The same assumptions apply to the "bias" corrected version of $Q_{a,\Omega,\lambda}(g;0)$ which is defined in Section 3.

ASSUMPTION 2.6 (Conditions on the spatial process). (a) There exists a $\delta > 0$, where $|c(\boldsymbol{s})| \leq \beta_{2+\delta}(\boldsymbol{s})$.

Required for the bounded frequency grid, to obtain the covariance of $\{J_n(\boldsymbol{\omega}_{\Omega,k})\}_k$.

(b) There exists a $\delta > 0$, where $f(\boldsymbol{\omega}) \leq \beta_{1+\delta}(\boldsymbol{\omega})$.

Required for the unbounded frequency grid—using this assumption instead of (a) in the case of bounded frequency grids leads to slightly larger errors bounds in the derivation of the mean and variance of $Q_{a,\Omega,\lambda}(g;r)$. This assumption is also used to obtain the CLT result for both the bounded and unbounded frequency grids.

(c) For all $1 \leq j \leq d$ and some $\delta > 0$, the partial derivatives satisfy $|\frac{\partial f(\boldsymbol{\omega})}{\partial \omega_j}| \leq \beta_{1+\delta}(\boldsymbol{\omega})$.

We use this condition to approximate sums with integral for both the bounded and unbounded frequency grids. It is also used to make a series of approximations to derive the limiting variance of $Q_{a,\Omega,\lambda}(g;r)$ in the case that the frequency grid is unbounded.

(d) There exists a $\delta > 0$, where $|\frac{\partial^d f(\boldsymbol{\omega})}{\partial \omega_1, \dots, \partial \omega_d}| \leq \beta_{1+\delta}(\boldsymbol{\omega})$.
Required only in the proof of Theorem 4.1(ii)(b).
(e) There exists a $\delta > 0$, where $|f(\boldsymbol{\omega})| \leq \beta_{2+\delta}(\boldsymbol{\omega})$.
Required only for the fixed-domain asymptotics.

REMARK 2.2.   Assumption 2.6(a) is satisfied by a wide range of covariance functions. Examples include:

 (i) The Wendland covariance, since its covariance is bounded and has a compact support.

(ii) The Matern covariance, which for $\nu > 0$ is defined as $c_\nu(\|\boldsymbol{s}\|_2) = \|\boldsymbol{s}\|_2^\nu \times K_\nu(\|\boldsymbol{s}\|_2)$ (K_ν is the modified Bessel function of the second kind); see [34]. To see why, we note that if $\nu > 0$ then $c_\nu(\boldsymbol{s})$ is a bounded function. Furthermore, for large $\|\boldsymbol{s}\|_2$, $c_\nu(\|\boldsymbol{s}\|_2) \sim C_\nu \|\boldsymbol{s}\|_2^{\nu-0.5} \exp(-\|\boldsymbol{s}\|_2)$ as $\|\boldsymbol{s}\|_2 \to \infty$ (where C_ν is a finite constant). Thus, by using the inequality

$$d^{-1/2} (|s_1| + |s_2| + \cdots + |s_d|) \leq \sqrt{s_1^2 + s_2^2 + \cdots + s_d^2} \leq (|s_1| + |s_2| + \cdots + |s_d|)$$

we can show $|c_\nu(\boldsymbol{s})| \leq \beta_{2+\delta}(\boldsymbol{s})$ for any $\delta > 0$.

REMARK 2.3.   Assumption 2.6(b,c,d) appears quite technical, but it is satisfied by a wide range of spatial covariance functions. For example, the spectral density of the Matern covariance defined in Remark 2.2 is $f_\nu(\boldsymbol{\omega}) = \frac{2^{\nu-1}\Gamma(\nu+\frac{d}{2})}{\pi^{d/2}(1+\|\boldsymbol{\omega}\|_2^2)^{(\nu+d/2)}}$ (see [34], p. 49). It is straightforward to show that this spectral density satisfies Assumption 2.6(b,c,d), noting that the δ used to define $\beta_{1+\delta}$ will vary with ν, dimension d and order of derivative.

If the spatial random field is non-Gaussian, we require the following assumptions on the higher order cumulants.

ASSUMPTION 2.7 (Non-Gaussian random fields). $\{Z(\boldsymbol{s}); \boldsymbol{s} \in \mathbb{R}^d\}$ is a fourth-order stationary spatial random field, in the sense that $E[Z(\boldsymbol{s})] = 0$, $\text{cov}[Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)] = c(\boldsymbol{s}_1 - \boldsymbol{s}_2)$, $\text{cum}[Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2), Z(\boldsymbol{s}_3)] = \kappa_2(\boldsymbol{s}_1 - \boldsymbol{s}_2, \boldsymbol{s}_1 - \boldsymbol{s}_3)$ and $\text{cum}[Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2), Z(\boldsymbol{s}_3), Z(\boldsymbol{s}_4)] = \kappa_4(\boldsymbol{s}_1 - \boldsymbol{s}_2, \boldsymbol{s}_1 - \boldsymbol{s}_3, \boldsymbol{s}_1 - \boldsymbol{s}_4)$, for some functions $\kappa_3(\cdot)$ and $\kappa_4(\cdot)$ and all $\boldsymbol{s}_1, \dots, \boldsymbol{s}_4 \in \mathbb{R}^d$. We define the fourth-order spectral density as $f_4(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \boldsymbol{\omega}_3) = \int_{\mathbb{R}^{3d}} \kappa_4(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3) \exp(-i \sum_{j=1}^3 \boldsymbol{s}_j' \boldsymbol{\omega}_j) d\omega_1 d\omega_2 d\omega_3$. We assume that for some $\delta > 0$ the spatial tri-spectral density function is such that $|f_4(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \boldsymbol{\omega}_3)| \leq \beta_{1+\delta}(\boldsymbol{\omega}_1)\beta_{1+\delta}(\boldsymbol{\omega}_2)\beta_{1+\delta}(\boldsymbol{\omega}_3)$ and $|\frac{\partial f_4(\omega_1,\dots,\omega_{3d})}{\partial \omega_j}| \leq \beta_{1+\delta}(\boldsymbol{\omega}_1)\beta_{1+\delta}(\boldsymbol{\omega}_2)\beta_{1+\delta}(\boldsymbol{\omega}_3)$.

2.2. *Properties of Fourier transforms.* In this section we briefly summarize some of the characteristics of the Fourier transforms $J_n(\omega_{\Omega,k})$. These results will be used in the construction of several estimators.

THEOREM 2.1 (Increasing domain asymptotics). *Let us suppose that* $\{Z(s);$ $s \in \mathbb{R}^d\}$ *is a stationary spatial random field whose covariance function [defined in Assumption* 2.1(i)] *satisfies Assumption* 2.6(a) *for some* $\delta > 0$. *Furthermore, the locations* $\{s_j\}$ *satisfy Assumption* 2.2. *Then we have*

$$(2.3) \quad \text{cov}\big[J_n(\omega_{k_1}), J_n(\omega_{k_2})\big] = \langle \gamma, \gamma_{(k_2-k_1)}\rangle f(\omega_{k_1}) + \frac{c(0)\gamma_{k_2-k_1}\lambda^d}{n} + O\left(\frac{1}{\lambda}\right),$$

where the bounds are uniform in $k_1, k_2 \in \mathbb{Z}^d$ *and* $\langle \gamma, \gamma_r \rangle = \sum_{j \in \mathbb{Z}^d} \gamma_j \gamma_{r-j}$.

Further, suppose Assumption 2.6(c) *holds and that* $|\gamma_j| \le \prod_{i=1}^d \xi_{2+\delta}(j_i)$ *for some* $\delta > 0$. *If* $\Omega > \lambda$ (*fine frequency grid is used*), *then*

$$(2.4) \quad \begin{aligned} &\text{cov}\big[J_n(\omega_{\Omega,k_1}), J_n(\omega_{\Omega,k_2})\big] \\ &\quad = f(\omega_{\Omega,k_1}) \sum_{j_1, j_2 \in \mathbb{Z}^d} \gamma_{j_1}\gamma_{j_2} \text{Sinc}\left(\pi\left[(j_1 + j_2) - \frac{\lambda}{\Omega}(k_1 - k_2)\right]\right) \\ &\qquad + O\left(\frac{\lambda^d}{n} + \frac{\log \lambda}{\lambda}\right). \end{aligned}$$

PROOF. See [36], Appendix A. □

Comparing (2.3) with (2.4), we see that if k_1 and k_2 are such that $\frac{\lambda}{2} \times \max |\omega_{\Omega,k_1} - \omega_{\Omega,k_2}| < 1$ then there is a high amount of correlation between the Fourier transforms. On the other hand, if $\lambda \max |\omega_{\Omega,k_1} - \omega_{\Omega,k_2}| \to \infty$ as $\lambda \to \infty$, the correlation declines. This means if the frequency grid is very coarse, $\Omega \ll \lambda$ (as considered in [1] and [2]) the DFTs are almost uncorrelated. On the other hand, if the frequency grid is very fine, $\Omega \gg \lambda$ [see (2.4)] frequencies which are close to each other are highly correlated. These observations suggest that estimators based on $J_n(\omega_{\Omega,k})$ do not gain in efficiency when $\Omega > \lambda$ but there will be a loss in efficiency when $\Omega < \lambda$. We show this heuristic to be true in Section 3.

The results in the above theorem give the limit within the increasing domain framework. In Theorem G.1, [36], we obtain the properties of $J_n(\omega_{\Omega,k})$ within the fixed-domain framework, where λ is kept fixed but $n \to \infty$. The expressions are long, but we summarize the most relevant parts in the remark below.

REMARK 2.4 (Fixed-domain asymptotics). (i) Let

$$(2.5) \quad A_\lambda\left(\frac{k}{\Omega}\right) = \int_{[-\lambda,\lambda]^d} T\left(\frac{u}{\lambda}\right) c(u) \exp\left(\frac{2i\pi k' u}{\Omega}\right) du,$$

where $T(\cdot)$ is the d-dimensional triangle kernel. Using Theorem G.1(iii), [36], under Assumption 2.3 (locations are uniformly distributed) we have $\mathrm{var}[J_n(\frac{2\pi k}{\Omega})] = A_\lambda(\frac{k}{\Omega}) + \frac{\lambda^d c(0)}{n}$.

Therefore, if the sampling frequency, Ω, is chosen such that $\Omega \geq 2\lambda$, then $\frac{1}{\Omega^d} A_\lambda(\frac{k}{\Omega})$ are the Fourier coefficients of $T(\frac{u}{\lambda})c(u)$ defined on the domain $[-\Omega/2, \Omega/2]^d$. In Section 2.3, we show that this fixed-domain approximation can, in some cases, be used to obtain unbiased estimators.

(ii) If the locations are not uniformly distributed, then by keeping λ fixed and using Theorem G.1(i) we see that $\mathrm{var}[J_n(\frac{2\pi k}{\Omega})]$ is not a separable function of the spatial spectral density and the Fourier coefficients of the random design. Whereas the approximation of $\mathrm{var}[J_n(\frac{2\pi k}{\Omega})]$ as $\lambda \to \infty$ given in Theorem 2.1 is a separable function of the spatial spectral density and spatial design. Using the separable approximation as the basis of an estimation scheme is simpler than using the exact but nonseparable formula.

REMARK 2.5. We observe from Theorems 2.1 that in the increasing domain framework $|J_n(\omega_{\Omega,k})|^2$ is an estimator of the spectral density function, $f(\omega_{\Omega,k})$ whereas within the fixed-domain framework $|J_n(\omega_{\Omega,k})|^2$ is an estimator of the Fourier coefficient $A_\lambda(k/\Omega)$. However, if $\int_{\mathbb{R}^d} |f(\omega)| \, d\omega < \infty$ and the ratio k/Ω is kept fixed then $A_\lambda(\frac{k}{\Omega}) \to f(\omega_{\Omega,k})$ as $\lambda \to \infty$.

2.3. *Examples of estimators defined within the Fourier domain.* Many parameters or quantities of interest can be written as a linear functional involving the spectral density function f. In Theorem 2.1 and Remark 2.5, we showed that if λ is large and the design of locations uniform then $\mathrm{E}[J_n(\omega_{\Omega,k})] \approx f(\omega_{\Omega,k})$ (if the design is not uniform then there will be an additional multiplicative constant). Motivated by this observation, in this section we consider estimators (or criterions) which take the form (1.5). If the locations follow a uniform distribution, then for some of the examples below it is possible to reduce the (fixed-domain) bias in the estimator.

2.3.1. *The Whittle likelihood.* Suppose the stationary spatial process $\{Z(s); s \in \mathbb{R}^d\}$ has spectral density $f_{\theta_0}(\omega)$ [and corresponding covariance $c_{\theta_0}(s)$] where θ_0 is unknown but belongs to the compact parameter space Θ. Matsuda and Yajima [27] propose using the integrated Whittle likelihood to estimate θ_0. More precisely, they define the Whittle likelihood as

$$\mathcal{L}_{I,n}(\theta, \eta^2) = \int_\Omega \left(\log[f_\theta(\omega) + \eta^2] + \frac{|J_n(\omega)|^2}{[f_\theta(\omega) + \eta^2]} \right) d\omega,$$

and use $(\widehat{\theta}, \widehat{\eta}) \in \arg\min_{\theta,\eta} \mathcal{L}_n(\theta, \eta)$ as an estimator of θ and η (where η is an estimator of the "ridge effect"). Of course, this integral cannot be evaluated in

practice and a Riemann sum approximation is necessary. Using $\{\boldsymbol{\omega_k} = 2\pi \boldsymbol{k}/\lambda; \boldsymbol{k} = (k_1, \ldots, k_d), -C\lambda \leq k_j \leq C\lambda\}$, we approximate the integral with the sum

$$\mathcal{L}_{S,n}(\theta, \eta^2) = \frac{1}{\lambda^d} \sum_{\boldsymbol{k}=-a}^{a} \left(\log[f_\theta(\boldsymbol{\omega_k}) + \eta^2] + \frac{|J_n(\boldsymbol{\omega_k})|^2}{[f_\theta(\boldsymbol{\omega_k}) + \eta^2]} \right).$$

A heuristic motivation for the above likelihood is that in the case the locations are uniformly distributed then $\{J_n(\boldsymbol{\omega_k})\}$ are near uncorrelated random variables with asymptotic variance $f_{\theta_0}(\boldsymbol{\omega_k}) + \eta_0^2$. If $J_n(\boldsymbol{\omega_k})$ were Gaussian, uncorrelated random variables with variance $f_{\theta_0}(\boldsymbol{\omega_k}) + \eta_0^2$ then $\mathcal{L}_{S,n}(\theta, \eta^2)$ would be the true likelihood. The choice of $a = C\lambda$ is necessary (where C does not depend on λ), since $|f_\theta(\boldsymbol{\omega})| \to 0$ as $\|\boldsymbol{\omega}\| \to \infty$ (for any norm $\|\cdot\|$), thus the discretized Whittle likelihood is only well defined over a bounded frequency grid. The choice of C is tied to how fast the tails in the parametric class of spectral density functions $\{f_\theta; \theta \in \Theta\}$ decay to zero.

If either Assumption 2.3 or 2.4 is satisfied, then we observe from Remark 2.4 that $A_\lambda(\frac{\boldsymbol{k}}{\lambda}; \theta_0)$ is a better approximation of $\text{var}[J_n(\boldsymbol{\omega_k})]$ than $f_{\theta_0}(\boldsymbol{\omega_k})$ [where $A_\lambda(\cdot)$ is defined in (2.5)]. Therefore, if Assumption 2.3 or 2.4 is satisfied, a better finite sample approximation can be obtained by using $\widehat{\theta} = \arg\min \mathcal{L}_{S,n}(\theta)$ as an estimator of θ, where

$$(2.6) \qquad \mathcal{L}_{S,n}(\theta, \eta^2) = \frac{1}{\lambda^d} \sum_{\boldsymbol{k}=-a}^{a} \left(\log\left[A_\lambda\left(\frac{\boldsymbol{k}}{\lambda}; \theta\right) + \eta^2 \right] + \frac{|J_n(\boldsymbol{\omega_k})|^2}{[A_\lambda(\frac{\boldsymbol{k}}{\lambda}; \theta) + \eta^2]} \right)$$

and $A_\lambda(\frac{\boldsymbol{k}}{\lambda}; \theta) = \int_{[-\lambda,\lambda]^d} T(\frac{\boldsymbol{u}}{\lambda}) c_\theta(\boldsymbol{u}) \exp(\frac{2i\pi \boldsymbol{k}'\boldsymbol{u}}{\lambda}) \, d\boldsymbol{u}$.

2.3.2. *The spectral density estimator.* We recall from Theorem 2.1 that $\text{var}[J_n(\boldsymbol{\omega_k})] \approx \langle \gamma, \gamma_0 \rangle f(\boldsymbol{\omega_k})$. Since $f(\cdot)$ is locally constant in a neighbourhood of $\boldsymbol{\omega}$ and motivated by spectral methods in time series, we use $\widehat{f}_{\lambda,n}(\boldsymbol{\omega})$ as a nonparametric estimator of f (or a constant multiple of it), where

$$\widehat{f}_{\lambda,n}(\boldsymbol{\omega}) = \sum_{\boldsymbol{k}=-\lambda/2}^{\lambda/2} W_b(\boldsymbol{\omega} - \boldsymbol{\omega_k}) |J_n(\boldsymbol{\omega_k})|^2 = \frac{1}{b^d} Q_{a,\lambda}(W_b, 0),$$

$W_b(\boldsymbol{\omega}) = b^{-d} \prod_{j=1}^{d} W(\frac{\omega_j}{b})$ and $W : [-1/2, -1/2] \to \mathbb{R}$ is a spectral window. In this case, we set the number of frequencies $a = \lambda/2$, and Assumption 2.5(i) is satisfied.

2.3.3. *A nonparametric nonnegative definite estimator of the spatial covariance.* In this section we propose a nonparametric estimator of the covariance. The estimator is based on the representation

$$c(\boldsymbol{u}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f(\boldsymbol{\omega}) \exp(i\boldsymbol{\omega}'\boldsymbol{u}) \, d\boldsymbol{\omega}.$$

Since the expectation of $|J_n(\boldsymbol{\omega_k})|^2$ is approximately $f(\boldsymbol{\omega_k})$, to estimate the spatial covariance we propose approximating the above integral with a sum and the spectral density with the absolute square of the Fourier transform. However, using the frequency grid $\boldsymbol{\omega_k} = \frac{2\pi\boldsymbol{k}}{\lambda}$ is problematic outside the region $[-\lambda/2, \lambda/2]^d$. Instead, we propose using a finer grid to estimate the covariance, namely

$$\widetilde{c}_{\Omega,n}(\boldsymbol{u}) = \frac{1}{\Omega^d} \sum_{\boldsymbol{k}=-a}^{a} |J_n(\boldsymbol{\omega}_{\Omega,\boldsymbol{k}})|^2 \exp(i\boldsymbol{u}'\boldsymbol{\omega}_{\Omega,\boldsymbol{k}}), \qquad \boldsymbol{u} \in \left[-\frac{\Omega}{2}, \frac{\Omega}{2}\right]^d,$$

where $\Omega \geq 2\lambda$. In this case, $a = a(\Omega)$ can be chosen such that $a/\Omega \to \infty$ as $a \to \infty$ and $\Omega \to \infty$ [thus Assumption 2.5(ii) is satisfied].

A disadvantage with the above "raw estimator" of the covariance is that there is no guarantee that it yields a nonnegative definite spatial auto-covariance function. However, this can easily be remedied by multiplication of $\widetilde{c}_{\Omega,n}(\boldsymbol{u})$ with the triangle kernel. More precisely, we define the estimator

$$\widehat{c}_{\Omega,n}(\boldsymbol{u}) = T\left(\frac{\boldsymbol{u}}{\widehat{\Omega}}\right)\widetilde{c}_{\Omega,n}(\boldsymbol{u}),$$

where $T(\boldsymbol{u}) = \prod_{j=1}^{d} T(u_j)$ and $\widehat{\Omega} \leq \Omega$. This covariance estimator has the advantage that it is zero outside the region $[-\widehat{\Omega}, \widehat{\Omega}]^d$. Moreover, $\widehat{c}_{\Omega,n}(\boldsymbol{u})$ is a nonnegative definite sequence. To show this result, we use that the Fourier transform of the triangle kernel, $T(u)$ is $\mathrm{sinc}^2(\frac{\omega}{2})$. Thus, the Fourier transform of $\widehat{c}_{\Omega,n}(\boldsymbol{u})$ is

$$\widehat{f}_\Omega(\boldsymbol{\omega}) = \int_{[-\widehat{\Omega},\widehat{\Omega}]^d} \widehat{c}_{\Omega,n}(\boldsymbol{u}) \exp(-i\boldsymbol{\omega}'\boldsymbol{u}) \, d\boldsymbol{u}$$

$$= \frac{\widehat{\Omega}^d}{\Omega^d} \sum_{\boldsymbol{k}=-a}^{a} |J_n(\boldsymbol{\omega}_{\Omega,\boldsymbol{k}})|^2 \, \mathrm{Sinc}^2\left[\frac{\widehat{\Omega}}{2}(\boldsymbol{\omega}_{\Omega,\boldsymbol{k}} - \boldsymbol{\omega})\right].$$

Clearly, $\widehat{f}_\Omega(\boldsymbol{\omega}) \geq 0$, therefore, the estimator $\{\widehat{c}_{\Omega,n}(\boldsymbol{u})\}$ is a nonnegative definite function, and thus a valid covariance function.

In Appendix J, [36], we illustrate the performance of the nonparametric nonnegative definite estimator of the spatial covariance with some simulations.

2.3.4. *A nonlinear least squares estimator of a parametric covariance function.* We recall that the Whittle likelihood can only be defined on a bounded frequency grid. This can be an issue if the observed locations are dense on the spatial domain, and thus contain a large amount of high frequency information which would be missed by the Whittle likelihood. An alternative method for parameter estimation of a spatial process is to use a different loss function. Motivated by [31], the discussion on the Whittle estimator in Section 2.3.1 and Theorem 2.1 we define the quadratic loss function

$$L_n(\theta) = \frac{1}{\lambda^d} \sum_{\boldsymbol{k}=-a}^{a} \left(|J_n(\boldsymbol{\omega_k})|^2 - \langle\gamma, \gamma_0\rangle f_\theta(\boldsymbol{\omega_k})\right)^2,$$

and let $\widehat{\theta}_n = \arg\min_{\theta \in \Theta} L_n(\theta)$ or equivalently solve $\nabla_\theta L_n(\theta) = 0$, where

$$\nabla_\theta L_n(\theta) = -\frac{2\langle \gamma, \gamma_0 \rangle}{\lambda^d} \sum_{k=-a}^{a} \nabla_\theta f_\theta(\boldsymbol{\omega_k})\{|J_n(\boldsymbol{\omega_k})|^2 - \langle \gamma, \gamma_0 \rangle f_\theta(\boldsymbol{\omega_k})\}$$

$$= -\langle \gamma, \gamma_0 \rangle \left[Q_{a,\lambda,\lambda}(2\nabla_\theta f_\theta(\cdot); 0) - \langle \gamma, \gamma_0 \rangle \frac{2}{\lambda^d} \sum_{k=-a}^{a} f_\theta(\boldsymbol{\omega_k})\nabla_\theta f_\theta(\boldsymbol{\omega_k}) \right].$$

It is well known that the distributional properties of a quadratic loss function are determined by its first derivative. In particular, the asymptotic sampling properties of $\widehat{\theta}_n$ are determined by $Q_{a,\lambda,\lambda}(2\nabla_\theta f(\cdot; \theta); 0)$. In this case, a can be such that $a/\lambda \to \infty$ as $\lambda \to \infty$ and Assumption 2.5(ii) is satisfied. An estimator of $\langle \gamma, \gamma_0 \rangle$ is given in Remark 4.1. Note that in the definition of $L_n(\theta)$, $\langle \gamma, \gamma_0 \rangle$ can be replaced with σ^2, in which case one is estimating a multiple of $f_{\theta_0}(\boldsymbol{\omega})$.

If either Assumption 2.3 or 2.4 is satisfied, then we can replace $f_\theta(\boldsymbol{\omega_k})$ with $A_\lambda(\frac{\boldsymbol{k}}{\lambda}; \theta)$, to obtain a better fixed-domain approximation.

3. A summary of the sampling properties of $Q_{a,\Omega,\lambda}(g; 0)$. In this section we consider the sampling properties of $Q_{a,\Omega,\lambda}(g; 0)$ for the general frequency grid $\{\boldsymbol{\omega}_{\Omega,k} = \frac{2\pi \boldsymbol{k}}{\Omega}\}$. The proof and more general results can be found in Appendix D, [36]. To simplify notation in this section we mainly consider the case that the locations are uniformly distributed.

LEMMA 3.1. *Suppose Assumptions* 2.1(i), 2.3 *and* 2.6(a,c) *or* (b,c) *hold. Let* $I(g; \frac{a}{\Omega})$ *and* $A_\lambda(\cdot)$ *be defined as in* (1.3) *and* (2.5), *respectively. Then*

$$(3.1) \quad E[Q_{a,\Omega,\lambda}(g; 0)] = \frac{c_2}{\Omega^d} \sum_{k=-a}^{a} g\left(\frac{2\pi \boldsymbol{k}}{\Omega}\right) A_\lambda\left(\frac{\boldsymbol{k}}{\Omega}\right) + \frac{c(0)\lambda^d}{n\Omega^d} \sum_{k=-a}^{a} g(\boldsymbol{\omega}_{\Omega,k}),$$

where $c_2 = n(n-1)/2$. *If we let* $\lambda \to \infty$, *then*

$$(3.2) \quad E[Q_{a,\Omega,\lambda}(g; 0)] = c_2 I\left(g; \frac{a}{\Omega}\right) + \frac{c(0)\lambda^d}{n\Omega^d} \sum_{k=-a}^{a} g(\boldsymbol{\omega}_{\Omega,k}) + O\left(\frac{\log \lambda}{\lambda} + \frac{1}{\Omega}\right).$$

PROOF. See [36], Appendix A. □

We observe an exact expression for the expectation of $Q_{a,\Omega,\lambda}(g; 0)$ is in terms of the Fourier coefficients $A_\lambda(\boldsymbol{k}/\Omega)$. However, an approximation of the expection of $Q_{a,\Omega,\lambda}(g; 0)$ (within the increasing domain framework) is in terms of an integral of the spectral density function.

We apply the above results to some of the examples considered in the previous section. The results are given in the general case that the locations are random variables but not necessarily uniformly distributed (see Theorem 4.2 and Lemma A.2 for the details).

EXAMPLE 3.1.   (i) *The Whittle likelihood.* Under Assumption 2.2, using Theorems 2.1 and 4.2, within the increasing domain asymptotics framework we have

$$
\begin{aligned}
& E[\mathcal{L}_{S,n}(\theta, \eta^2)] \\
& = \frac{1}{(2\pi)^d} \int_{[-a/\lambda, a/\lambda]^d} \left(\log[f_\theta(\boldsymbol{\omega}) + \eta^2] + \frac{\langle \gamma, \gamma_0 \rangle f_{\theta_0}(\boldsymbol{\omega}) + \gamma_0 \eta_0^2}{f_\theta(\boldsymbol{\omega}) + \eta^2} \right) d\omega \\
& \quad + O\left(\frac{1}{\lambda}\right),
\end{aligned}
$$

where $f_{\theta_0}(\cdot)$ denotes the true spectral density, $\eta_0^2 = \lambda^d n^{-1} c(0; \theta_0)$ [note that $\eta_0^2 = O(\lambda^d/n)$] with $c(\boldsymbol{s}; \theta_0)$ the corresponding spatial covariance.

Assuming Assumption 2.3 holds, within the fixed-domain framework the expectation of (2.6) is

$$
\begin{aligned}
E[\mathcal{L}_{S,n}(\theta, \eta^2)] &= \frac{1}{\lambda^d} \sum_{\boldsymbol{k}=-a}^{a} \left(\log\left[A_\lambda\left(\frac{\boldsymbol{k}}{\lambda}; \theta\right) + \eta^2 \right] + \frac{A_\lambda(\frac{\boldsymbol{k}}{\lambda}; \theta_0)}{[A_\lambda(\frac{\boldsymbol{k}}{\lambda}; \theta) + \eta^2]} \right) \\
&\quad + O\left(\frac{1}{n}\left(\frac{a}{\lambda}\right)^{2d}\right),
\end{aligned}
$$

where in the above error bound we use that the tails of $A_\lambda(\boldsymbol{k}/\lambda; \theta)$ decay at the rate $\prod_{i=1}^{d} \xi_2(k_i)$.

(ii) *The nonparametric covariance.* Within the increasing domain framework and using Lemma A.2(ii), [36], for $\boldsymbol{u} \in [-\min(\lambda, \Omega/2), \min(\lambda, \Omega/2)]^d$ we have

$$
E[\widetilde{c}_{\Omega,n}(\boldsymbol{u})] = \langle \gamma, \gamma_0 \rangle c(\boldsymbol{u}) + O\left(\frac{\log \lambda}{\lambda} + \frac{1}{\Omega}\right),
$$

and $E[\widehat{c}_{\Omega,n}(\boldsymbol{u})] = \langle \gamma, \gamma_0 \rangle T(\frac{\boldsymbol{u}}{\Omega}) c(\boldsymbol{u}) + O(T(\frac{\boldsymbol{u}}{\Omega})[\frac{\log \lambda}{\lambda} + \frac{1}{\Omega}])$ where $\langle \gamma, \gamma_0 \rangle$ is defined in Theorem 2.1.

In order to understand the properties of $\widetilde{c}_{\Omega,n}(\boldsymbol{u})$ within the fixed-domain framework, we assume that Assumption 2.3 holds. If $\Omega \geq 2\lambda$ and $\boldsymbol{u} \in [-\lambda, \lambda]$, then by using Lemma 3.1 we have

$$
\begin{aligned}
E[\widetilde{c}_{\Omega,n}(\boldsymbol{u})] &= \frac{1}{\Omega^d} \sum_{\boldsymbol{k}=-\infty}^{\infty} A_\lambda\left(\frac{\boldsymbol{k}}{\Omega}\right) \exp(i\boldsymbol{u}'\boldsymbol{\omega}_{\Omega,k}) + O\left(\frac{1}{n} + \frac{1}{a}\right) \\
&= T\left(\frac{\boldsymbol{u}}{\lambda}\right) c(\boldsymbol{u}) + O\left(\frac{1}{n} + \frac{1}{a}\right),
\end{aligned}
$$

where we recall $T(\cdot)$ denotes the triangle kernel.

In Lemma 3.1, we observe the bias $\frac{c(0)\lambda^d}{n\Omega^d}\sum_{k=-a}^{a}g(\boldsymbol{\omega}_{\Omega,k})$. It can be removed by using a bias corrected version of $Q_{a,\Omega,\lambda}(g;0)$

(3.3)
$$\widetilde{Q}_{a,\Omega,\lambda}(g;0) = \frac{1}{\Omega^d}\sum_{k=-a}^{a}g(\boldsymbol{\omega}_{\Omega,k})\left|J_n(\boldsymbol{\omega}_{\Omega,k})\right|^2$$
$$- \frac{\lambda^d}{\Omega^d n}\sum_{k=-a}^{a}g(\boldsymbol{\omega}_{\Omega,k})\frac{1}{n}\sum_{j=1}^{n}Z(\boldsymbol{s}_j)^2.$$

For the remainder of this section, we focus on this bias corrected estimator. The analogous result for $Q_{a,\Omega,\lambda}(g;0)$ can be found in Appendix H, [36].

We show in Appendix D, [36] that in the case that $\{Z(\boldsymbol{u}); \boldsymbol{u} \in \mathbb{R}^d\}$ is a Gaussian stationary spatial process

$$\mathrm{var}[\widetilde{Q}_{a,\Omega,\lambda}(g;0)] \approx C_1\left(\frac{a}{\Omega}\right)\frac{1}{\Omega^d}\sum_{k_1,\ldots,k_d=-2a}^{2a}\mathrm{Sinc}^2\left(\frac{\lambda}{\Omega}k\pi\right),$$

where

(3.4) $$C_1\left(\frac{a}{\Omega}\right) = \frac{1}{(2\pi)^d}\int_{-[2\pi a/\Omega,2\pi a/\Omega]^d}f(\boldsymbol{\omega})^2\left[|g(\boldsymbol{\omega})|^2 + g(\boldsymbol{\omega})\overline{g(-\boldsymbol{\omega})}\right]d\boldsymbol{\omega}.$$

Observe that the rate of convergence of $\mathrm{var}[\widetilde{Q}_{a,\Omega,\lambda}(g;0)]$ is determined by

$$\sum_{k_1,\ldots,k_d=-2a}^{2a}\mathrm{Sinc}^2\left(\frac{\lambda}{\Omega}k\pi\right) = \prod_{i=1}^{d}\sum_{k_i=-2a}^{2a}\mathrm{sinc}^2\left(\frac{\lambda}{\Omega}k_i\pi\right).$$

It is this term along with the following result which gives the crucial insight into the rate of convergence for different frequency grids $\{\boldsymbol{\omega}_{\Omega,k}\}$. If $a \to \infty$, then

$$\frac{1}{\Omega}\sum_{k=-\infty}^{\infty}\mathrm{sinc}^2\left(\frac{\lambda}{\Omega}k\pi\right) = \begin{cases} \dfrac{1}{\lambda}, & \dfrac{\lambda}{\Omega} < 1, \\[2mm] \dfrac{1}{\Omega}, & \dfrac{\lambda}{\Omega} \in \mathbb{Z}, \\[2mm] O\left(\dfrac{1}{\Omega}\right), & \dfrac{\lambda}{\Omega} > 1 \text{ and } \dfrac{\lambda}{\Omega} \notin \mathbb{Z}, \end{cases}$$

further, if $\lambda/\Omega \to \infty$ then $\sum_{k=-\infty}^{\infty}\mathrm{sinc}^2(\frac{\lambda}{\Omega}k\pi) \to 1$ (see Appendix D, [36] for the proof). This result implies that

$$\mathrm{var}[\widetilde{Q}_{a,\Omega,\lambda}(g;0)] = \begin{cases} O\left(\dfrac{1}{\lambda^d}\right), & \lambda < \Omega, \\[2mm] O\left(\dfrac{1}{\Omega^d}\right), & \lambda \geq \Omega. \end{cases}$$

In other words, the frequency grid $\omega_{\lambda,k} = \frac{2\pi k}{\lambda}$ or finer will yield a rate of convergence of $O(\lambda^{-d})$ and $\mathrm{var}[\widetilde{Q}_{a,\Omega,\lambda}(g;0)] \approx \lambda^{-d}C_1(\frac{a}{\Omega})$. However, a coarse frequency grid $\omega_{\Omega,k} = \frac{2\pi k}{\Omega}$ where $\Omega < \lambda$ will yield a slower rate of convergence of $O(\Omega^{-d})$. In the theorem below, we make this precise.

For the following theorem, we consider general stationary spatial random fields. This requires the following definition:

$$
(3.5) \qquad
\begin{aligned}
D_1\left(\frac{a}{\Omega}\right) &= \frac{1}{(2\pi)^{2d}} \int_{[-2\pi a/\Omega,\, 2\pi a/\Omega]^{2d}} g(\boldsymbol{\omega}_1)\overline{g(\boldsymbol{\omega}_2)} \\
&\quad \times f_4(-\boldsymbol{\omega}_1, -\boldsymbol{\omega}_2, \boldsymbol{\omega}_2)\, d\boldsymbol{\omega}_1\, d\boldsymbol{\omega}_2.
\end{aligned}
$$

This term arises if the spatial random field is non-Gaussian.

THEOREM 3.1. *Suppose Assumptions* 2.1(i), 2.2, 2.5(i) *or* (ii), 2.6(b,c) *and* 2.7 *hold. Let* $C_1(\cdot)$ *and* $D_1(\cdot)$ *be defined as in* (3.4) *and* (3.5), *respectively*:

(i) *If a fine frequency grid is used* ($\frac{\lambda}{\Omega} < 1$), *then*

$$
\lambda^d\, \mathrm{var}[\widetilde{Q}_{a,\Omega,\lambda}(g;0)] = C_1\left(\frac{a}{\Omega}\right)\left[\frac{\lambda^d}{\Omega^d}\sum_{k=-2a}^{2a}\mathrm{Sinc}^2\left(\frac{\lambda}{\Omega}k\pi\right)\right] + D_1\left(\frac{a}{\Omega}\right) \\
+ O(\ell_{a,\Omega,\lambda}^{(2)}),
$$

(ii) *If a coarse frequency grid is used* ($\frac{\lambda}{\Omega} \geq 1$), *then*

$$
\Omega^d\, \mathrm{var}[\widetilde{Q}_{a,\Omega,\lambda}(g;0)] = C_1\left(\frac{a}{\Omega}\right)\left[\sum_{k=-2a}^{2a}\mathrm{Sinc}^2\left(\frac{\lambda}{\Omega}k\pi\right)\right] + \left(\frac{\Omega}{\lambda}\right)^d D_1\left(\frac{a}{\Omega}\right) \\
+ O(\widetilde{\ell}_{a,\Omega,\lambda}^{(2)}),
$$

where $\widetilde{\ell}_{a,\Omega,\lambda}^{(2)}$ *is defined in* (D.16), *Appendix D,* [36].

PROOF.   See [36], Appendix D.   □

As mentioned above, one important implication of the above result is that the rate of convergence depends on whether the frequency grid is coarser or finer than $1/\lambda$, where λ is the length of the spatial domain. In terms of the asymptotic sampling properties (see Lemma 3.1 and Theorem 3.1), there seems to be little benefit in using a very fine frequency grid, as it does not reduce the bias or variance (but is computationally costly).

We observe that if the spatial process is non-Gaussian then an additional term, $D_1(a/\Omega)$, involving the fourth-order cumulant of the spatial process, arises. However, if the frequency grid is extremely coarse in the sense that $\lambda/\Omega \to \infty$ as $\lambda \to \infty$ and $\Omega \to \infty$, then $D_1(a/\Omega)$ is asymptotic negligible compared with the

leading term which is a function of the spectral density. For example, if $\lambda/\Omega \in \mathbb{Z}^+$ then

$$\Omega^d \operatorname{var}[\widetilde{Q}_{a,\Omega,\lambda}(g;0)] = C_1\left(\frac{a}{\Omega}\right) + O\left(\ell_{a,\Omega,\lambda}^{(2)} + \frac{\Omega^d}{\lambda^d}\right).$$

Thus, if $\lambda/\Omega \to \infty$ as $\lambda \to \infty$ and $\Omega \to \infty$ [and $\overline{g(\omega)} = g(-\omega)$] we have verified condition (C.4) in [2],

$$\frac{\operatorname{var}[\sum_{k=-a}^{a} g(\omega_{\Omega,k})|J_n(\omega_{\Omega,k})|^2]}{2\sum_{k=-a}^{a} |g(\omega_{\Omega,k})|^2 f(\omega_{\Omega,k})^2} \xrightarrow{\mathcal{P}} 1,$$

which is required for their proposed spatial spectral empirical likelihood methodology. Therefore, a very coarse grid has the advantage that the term $D_1(\cdot)$ is negligible. However, we see from Lemma 3.1 and Theorem 3.1 that the disadvantage is that there is a substantial increase in both variance and bias.

Since the grid, $\omega_k = 2\pi k/\lambda$, yields optimal sampling properties in Section 4 we focus on deriving sampling properties of $\widetilde{Q}_{a,\lambda}(g;r)$, where

(3.6)
$$\begin{aligned}
\widetilde{Q}_{a,\lambda}(g;r) &= \frac{1}{\lambda^d} \sum_{k_1,\ldots,k_d=-a}^{a} g(\omega_k) J_n(\omega_k) \overline{J_n(\omega_{k+r})} \\
&\quad - \frac{1}{n} \sum_{k=-a}^{a} g(\omega_k) \frac{1}{n} \sum_{j=1}^{n} Z(s_j)^2 e^{-is_j'\omega_r}.
\end{aligned}$$

4. Sampling properties of $\widetilde{Q}_{a,\lambda}(g;r)$.

In this section we show that under the increasing domain framework $\widetilde{Q}_{a,\lambda}(g;r)$ [defined (3.6)] is a consistent estimator of $I(g;\frac{a}{\lambda})$ (or some multiple of it), where $I(g;\frac{a}{\lambda})$ is defined in (1.3). The sampling properties in the fixed-domain framework are given in Section 4.4.

4.1. *The expectation of $\widetilde{Q}_{a,\lambda}(g;r)$.* We start with the expectation of $\widetilde{Q}_{a,\lambda}(g;r)$. We show if $\sup_{\omega \in \mathbb{R}^d} |g(\omega)| < \infty$, the choice of a does not play a significant role in the asymptotic properties of $\widetilde{Q}_{a,\lambda}(g;r)$. If $a \gg \lambda$, the analysis of $\widetilde{Q}_{a,\lambda}(g;r)$ requires more delicate techniques. We start by stating some pertinent features in the analysis of $\widetilde{Q}_{a,\lambda}(g;r)$, which gives a flavour of our approach. By writing $\widetilde{Q}_{a,\lambda}(g;r)$ as a quadratic form, it is straightforward to show that

$$E[\widetilde{Q}_{a,\lambda}(g;r)]$$

(4.1)
$$\begin{aligned}
&= c_2 \sum_{k=-a}^{a} g(\omega_k) \frac{1}{\lambda^d} \int_{[-\lambda/2,\lambda/2]^{2d}} c(s_1 - s_2) \exp\left(i\omega_k'(s_1 - s_2) - is_2'\omega_r\right) \\
&\quad \times h\left(\frac{s_1}{\lambda}\right) h\left(\frac{s_2}{\lambda}\right) ds_1 \, ds_2,
\end{aligned}$$

where $c_2 = n(n-1)/n^2$. The proof of Theorem 2.1 is based on making a change of variables $v = s_1 - s_2$ and then systematically changing the limits of the integral. This method can be applied to the above, if a is such that the ratio a/λ is fixed for all λ. However, if the frequency grid $[-a/\lambda, a/\lambda]^d$ is allowed to grow with λ, applying this brute force method to $\mathrm{E}[\widetilde{Q}_{a,\lambda}(g; r)]$ has the disadvantage that it aggregates the errors within the sum of $\mathrm{E}[\widetilde{Q}_{a,\lambda}(g; r)]$. Instead, to further the analysis, we replace $c(s_1 - s_2)$ by its spectral representation $c(s_1 - s_2) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f(\omega) \exp(i\omega'(s_1 - s_2)) \, d\omega$ and focus on the case that the sampling design is uniform; $h(s/\lambda) = \lambda^{-d} I_{[-\lambda/2, \lambda/2]}(s)$ (later we consider general sampling densities). This reduces the first term in $\mathrm{E}[\widetilde{Q}_{a,\lambda}(g; r)]$ to the Fourier transforms of step functions, which is the product of sinc functions. Specifically, we obtain

$$\mathrm{E}[\widetilde{Q}_{a,\lambda}(g; r)]$$

$$= \frac{c_2}{(2\pi)^d} \sum_{k=-a}^{a} g(\omega_k) \int_{\mathbb{R}^d} f(\omega) \operatorname{Sinc}\left(\frac{\lambda\omega}{2} + k\pi\right) \operatorname{Sinc}\left(\frac{\lambda\omega}{2} + (k+r)\pi\right) d\omega$$

$$= \frac{c_2}{\pi^d} \int_{\mathbb{R}^d} \operatorname{Sinc}(y) \operatorname{Sinc}(y + r\pi) \left[\frac{1}{\lambda^d} \sum_{k=-a}^{a} g(\omega_k) f\left(\frac{2y}{\lambda} - \omega_k\right)\right] dy,$$

where the last line above is due to a change of variables $y = \frac{\lambda\omega}{2} + k\pi$. Since the spectral density function is absolutely integrable, it is clear that $[\frac{1}{\lambda^d} \sum_{k=-a}^{a} g(\omega_k) f(\frac{2y}{\lambda} - \omega_k)]$ is uniformly bounded over y and that $\mathrm{E}[\widetilde{Q}_{a,\lambda}(g; r)]$ is finite for all λ. Furthermore, if $f(\frac{2y}{\lambda} - \omega_k)$ were replaced with $f(-\omega_k)$, then what remains in the integral are two shifted sinc functions, which is zero if $r \in \mathbb{Z}^d/\{0\}$, that is,

$$\mathrm{E}[\widetilde{Q}_{a,\lambda}(g; r)] = \frac{c_2}{\pi^d} \int_{\mathbb{R}^d} \operatorname{Sinc}(y) \operatorname{Sinc}(y + r\pi) \left[\frac{1}{\lambda^d} \sum_{k=-a}^{a} g(\omega_k) f(-\omega_k)\right] dy + R,$$

where

$$R = \frac{c_2}{\pi^d} \int_{\mathbb{R}^d} \operatorname{Sinc}(y) \operatorname{Sinc}(y + r\pi)$$

$$\times \left[\frac{1}{\lambda^d} \sum_{k=-a}^{a} g(\omega_k) \left(f\left(\frac{2y}{\lambda} - \omega_k\right) - f(-\omega_k)\right)\right] dy.$$

In the following theorem, we show that under certain conditions on f, R is asymptotically negligible.

THEOREM 4.1.   *Let $I(g; \cdot)$ be defined as in* (1.3). *Throughout the theorem, we suppose Assumptions* 2.1(i) *and* 2.3 *hold. Let $b(r)$ denote the number of zero values in r:*

(i) *If Assumptions* 2.5(i) *and* 2.6(a,c) *hold, then we have*

(4.2)
$$
E[\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})] = \begin{cases} O\left(\dfrac{1}{\lambda^{d-b(\boldsymbol{r})}}\right), & \boldsymbol{r} \in \mathbf{Z}^d/\{\mathbf{0}\}, \\[2mm] I(g;C) + O\left(\dfrac{1}{\lambda}\right), & \boldsymbol{r} = \mathbf{0}. \end{cases}
$$

(ii) *Suppose Assumptions* 2.5(ii) *holds and*:

    (a) *Assumption* 2.6(b) *holds, then* $\sup_a |E[\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})]| < \infty$.

    (b) *Assumption* 2.6(b,c,d) *holds, then we have*

$$
E[\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})] = \begin{cases} O\left(\dfrac{1}{\lambda^{d-b(\boldsymbol{r})}} \displaystyle\prod_{j=1}^{d-b(\boldsymbol{r})} (\log \lambda + \log |m_j|)\right), & \boldsymbol{r} \in \mathbf{Z}^d/\{\mathbf{0}\}, \\[3mm] I\left(g;\dfrac{a}{\lambda}\right) + O\left(\dfrac{\log \lambda}{\lambda} + \dfrac{1}{n}\right), & \boldsymbol{r} = \mathbf{0}, \end{cases}
$$

where $\{m_1, \dots, m_{d-b(\boldsymbol{r})}\}$ *is the subset of nonzero values in* $\boldsymbol{r} = (r_1, \dots, r_d)$.

    (c) *If only Assumption* 2.6(b,c) *holds, then the* $O\left(\frac{1}{\lambda^{d-b(\boldsymbol{r})}} \prod_{j=1}^{d-b(\boldsymbol{r})} (\log \lambda + \log |m_j|)\right)$ *term in* (b) *is replaced with the slower rate* $O\left(\frac{1}{\lambda}(\log \lambda + \log[1 + \|\boldsymbol{r}\|_1])\right)$.

Note that the above bounds for (b) *and* (c) *are uniform in* a.

PROOF. See [36], Appendix A. □

We observe that if $\boldsymbol{r} \neq \mathbf{0}$, then $\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})$ is estimating zero as $\lambda \to \infty$. It would appear that $\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})$ when $\boldsymbol{r} \neq \mathbf{0}$ does not contain any useful information, however, in Section 5 we show how these terms can be used to estimate nuisance parameters.

In order to analyze $E[\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})]$ in the case that the locations are not from a uniform distribution, we return to (4.1) and replace $c(\boldsymbol{s}_1 - \boldsymbol{s}_2)$ and $h(\cdot)$ by their Fourier representations:

$$
E[\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})] = \frac{c_2}{\pi^d} \sum_{j_1,j_2 \in \mathbb{Z}} \gamma_{j_1} \gamma_{j_2} \frac{1}{\lambda^d} \sum_{k=-a}^{a} g(\boldsymbol{\omega_k}) \int_{-\infty}^{\infty} f\left(\frac{2\boldsymbol{y}}{\lambda} - \boldsymbol{\omega_k}\right)
$$
$$
\times \operatorname{Sinc}(\boldsymbol{y}) \operatorname{Sinc}(\boldsymbol{y} + (\boldsymbol{r} - j_1 - j_2)\pi) \, d\boldsymbol{y}.
$$

This representation allows us to use similar techniques to those used in the uniform sampling case to prove the following result.

THEOREM 4.2. *Let* $I(g;\cdot)$ *be defined as in* (1.3). *Suppose Assumptions* 2.1(i) *and* 2.2 *hold*:

(i) *If in addition Assumptions* 2.5(i) *and* 2.6(a,c) *hold, then we have*

$$E[\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})] = \langle \gamma, \gamma_{\boldsymbol{r}}\rangle I\left(g;\frac{a}{\lambda}\right) + O(\lambda^{-1}),$$

$O(\lambda^{-1})$ *is uniform over* $\boldsymbol{r} \in \mathbb{Z}^d$.

(ii) *If in addition Assumptions* 2.5(ii) *and* 2.6(b,c) *hold, then we have*

$$E[\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})] = \langle \gamma, \gamma_{-\boldsymbol{r}}\rangle I\left(g;\frac{a}{\lambda}\right) + O\left(\frac{\log \lambda + \log(1 + \|\boldsymbol{r}\|_1)}{\lambda}\right).$$

PROOF.   See [36], Appendix A.   □

We observe that by applying Theorem 4.2 to the case that h is uniform (using that $\gamma_0 = 1$ else $\gamma_j = 0$) gives $E[\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})] = O(\lambda^{-1})$ for $\boldsymbol{r} \neq 0$. Hence, in the case that the sampling is uniform, Theorems 4.1 and 4.2 give similar results, though the bounds in Theorem 4.1 are sharper.

REMARK 4.1 (Estimation of $\sum_{\boldsymbol{j} \in \mathbb{Z}^d} |\gamma_{\boldsymbol{j}}|^2$).   The above lemma implies that $E[\widetilde{Q}_{a,\lambda}(g;0)] = \langle \gamma, \gamma_0 \rangle I(g;\frac{a}{\lambda})$. Therefore, to estimate $I(g;\frac{a}{\lambda})$ we require an estimator of $\langle \gamma, \gamma_0 \rangle$. To do this, we recall that

$$\langle \gamma, \gamma_0 \rangle = \sum_{\boldsymbol{j} \in \mathbb{Z}} |\gamma_{\boldsymbol{j}}|^2 = \frac{1}{\lambda^d} \int_{[-\lambda/2, \lambda/2]^d} h\left(\frac{\boldsymbol{\omega}}{\lambda}\right)^2 d\boldsymbol{\omega}.$$

Therefore, one method for estimating the above integral is to define a grid on $[-\lambda/2, \lambda/2]^d$ and estimate $h(\cdot)$ at each point on the grid, then to take the average squared over the grid (see Remark 1, [27]). An alternative, computationally simpler method, is to use the method proposed in [18]. That is, use

$$\widehat{\langle \gamma, \gamma_0 \rangle} = \frac{2}{n(n-1)b} \sum_{1 \le j_1 < j_2 \le n} K\left(\frac{\boldsymbol{s}_{j_1} - \boldsymbol{s}_{j_2}}{b}\right)^2,$$

as an estimator of $\langle \gamma, \gamma_0 \rangle$, where $K : [-1/2, 1/2]^d \to \mathbb{R}$ is a kernel function. Note that multiplying the above kernel with $\exp(-i\boldsymbol{\omega}_{\boldsymbol{r}}' \boldsymbol{s}_{j_2})$ results in an estimator of $\langle \gamma, \gamma_{\boldsymbol{r}} \rangle$. In the case $d = 1$ and under certain regularity conditions, [18] show if the bandwidth b is selected in an appropriate way then $\widehat{\langle \gamma, \gamma_0 \rangle}$ attains the classical $O(n^{-1/2})$ rate under suitable regularity conditions (see, also, [5] and [25]). It seems plausible a similar result holds for $d > 1$ (though we do not prove it here). Therefore, an estimator of $I(g;\frac{a}{\lambda})$ is $\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})/\widehat{\langle \gamma, \gamma_0 \rangle}$.

4.2. *The covariance and asymptotic normality.*   In the previous section, we showed that the expectation of $\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})$ depends only on the number of frequencies a through the limit of the integral $I(g;\frac{a}{\lambda})$ [if $\sup_{\boldsymbol{\omega}\mathbb{R}^d} |g(\boldsymbol{\omega})| < \infty$]. In this section, we show that a plays a mild role in the higher order properties of $\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})$. We focus on the case that the random field is Gaussian and later describe how the results differ in the case that the random field is non-Gaussian.

THEOREM 4.3. *Suppose Assumptions* 2.1, 2.2 *hold. Let* $U_1(\boldsymbol{r}_1, \boldsymbol{r}_2; \boldsymbol{\omega}_{\boldsymbol{r}_1}, \boldsymbol{\omega}_{\boldsymbol{r}_2})$ *and* $U_2(\boldsymbol{r}_1, \boldsymbol{r}_2; \boldsymbol{\omega}_{\boldsymbol{r}_1}, \boldsymbol{\omega}_{\boldsymbol{r}_2})$ *be defined as in equation* (C.1), [36].

(i) *If Assumption* 2.5(i) *and* 2.6(a,c) *also hold. Then uniformly for all* $0 \leq \|\boldsymbol{r}_1\|_1, \|\boldsymbol{r}_2\|_1 \leq C|\lambda|$, *we have*

$$\lambda^d \operatorname{cov}\big[\widetilde{Q}_{a,\lambda}(g; \boldsymbol{r}_1), \widetilde{Q}_{a,\lambda}(g; \boldsymbol{r}_2)\big] = U_1(\boldsymbol{r}_1, \boldsymbol{r}_2; \boldsymbol{\omega}_{\boldsymbol{r}_1}, \boldsymbol{\omega}_{\boldsymbol{r}_2}) + O\left(\frac{1}{\lambda} + \frac{\lambda^d}{n}\right)$$

and

$$\lambda^d \operatorname{cov}\big[\widetilde{Q}_{a,\lambda}(g; \boldsymbol{r}_1), \overline{\widetilde{Q}_{a,\lambda}(g; \boldsymbol{r}_2)}\big] = U_2(\boldsymbol{r}_1, \boldsymbol{r}_2; \boldsymbol{\omega}_{\boldsymbol{r}_1}, \boldsymbol{\omega}_{\boldsymbol{r}_2}) + O\left(\frac{1}{\lambda} + \frac{\lambda^d}{n}\right).$$

(ii) *If Assumption* 2.5(ii) *and* 2.6(b) *also hold. Then*

$$\lambda^d \sup_{a, \boldsymbol{r}} \operatorname{var}\big[\widetilde{Q}_{a,\lambda}(g; \boldsymbol{r})\big] < \infty$$

with $\lambda^d/n \to c$ (*where* $0 \leq c < \infty$) *as* $\lambda \to \infty$ *and* $n \to \infty$.

(iii) *If Assumption* 2.5(ii) *and* 2.6(b,c) *also hold. Then uniformly for all* $0 \leq \|\boldsymbol{r}_1\|_1, \|\boldsymbol{r}_2\|_1 \leq C|a|$ (*for some finite constant* C) *we have*

$$\lambda^d \operatorname{cov}\big[\widetilde{Q}_{a,\lambda}(g; \boldsymbol{r}_1), \widetilde{Q}_{a,\lambda}(g; \boldsymbol{r}_2)\big] = U_1(\boldsymbol{r}_1, \boldsymbol{r}_2; \boldsymbol{\omega}_{\boldsymbol{r}_1}, \boldsymbol{\omega}_{\boldsymbol{r}_2}) + O(\ell_{\lambda,a,n}),$$

$$\lambda^d \operatorname{cov}\big[\widetilde{Q}_{a,\lambda}(g; \boldsymbol{r}_1), \overline{\widetilde{Q}_{a,\lambda}(g; \boldsymbol{r}_2)}\big] = U_2(\boldsymbol{r}_1, \boldsymbol{r}_2; \boldsymbol{\omega}_{\boldsymbol{r}_1}, \boldsymbol{\omega}_{\boldsymbol{r}_2}) + O(\ell_{\lambda,a,n}),$$

where

(4.3) $$\ell_{\lambda,a,n} = \log^2(a)\left[\frac{\log a + \log \lambda}{\lambda}\right] + \frac{\lambda^d}{n}.$$

PROOF. See [36], Appendix C. □

We now briefly discuss the above results. From Theorem 4.3(ii), we see that $\widetilde{Q}_{a,\lambda}(g; \boldsymbol{r})$ is a mean squared consistent estimator of $\langle \gamma, \gamma_{\boldsymbol{r}} \rangle I(g; \frac{a}{\lambda})$, that is, $\mathrm{E}[\widetilde{Q}_{a,\lambda}(g; \boldsymbol{r}) - \langle \gamma, \gamma_{\boldsymbol{r}} \rangle I(g; \frac{a}{\lambda})]^2 = O(\lambda^{-d} + (\frac{\log \lambda}{\lambda} + \frac{1}{n})^2)$ as $a \to \infty$ and $\lambda \to \infty$.

In order to obtain an explicit expression for the variance additional conditions are required. In particular, Theorem 4.3(iii) states that if the frequency grid is unbounded we require some additional conditions on the spectral density function and some mild constraints on the rate of growth of the frequency domain a. More precisely, a should be such that $a = O(\lambda^k)$ for some $1 \leq k < \infty$.

REMARK 4.2 (Selecting a in practice). The above gives theoretical guidelines. In practice, if $\sup_{\boldsymbol{\omega} \in \mathbb{R}^d} |g(\boldsymbol{\omega})| < \infty$ we suggest plotting $|J_n(\boldsymbol{\omega}_k)|^2$ against $\boldsymbol{\omega}_k$. $|J_n(\boldsymbol{\omega}_k)|^2$ will drop close to zero for large $\|\boldsymbol{\omega}_k\|_1$ (see Figure 1, Appendix J, [36]). Thus, a should be chosen such that it lies after this point. The precise value does not matter too much as the results are not too sensitive to the choice of a.

The expressions for $\text{cov}[\lambda^d \, \text{cov}[\widetilde{Q}_{a,\lambda}(g; r_1), \overline{\widetilde{Q}_{a,\lambda}(g; r_2)}]]$ (see equation (C.1), [36]) are unwieldy; however, some simplifications can be made if $\|r_1\|_1 \ll \lambda$ and $\|r_2\|_1 \ll \lambda$.

COROLLARY 4.1.   *Suppose Assumptions* 2.2, 2.5 *and* 2.6(a,c) *or* 2.6(b,c) *hold. Then*

(4.4)
$$U_1(r_1, r_2; \omega_{r_1}, \omega_{r_2}) = \Gamma_{r_1 - r_2} C_1 + O(\varepsilon_{r_1, r_2}(\lambda)),$$
$$U_2(r_1, r_2; \omega_{r_1}, \omega_{r_2}) = \Gamma_{r_1 + r_2} C_2 + O(\varepsilon_{r_1, r_2}(\lambda)),$$

where $\Gamma_r = \sum_{j_1 + j_2 + j_3 + j_4 = r} \gamma_{j_1} \gamma_{j_2} \gamma_{j_3} \gamma_{j_4}$, $\varepsilon_{r_1, r_2}(\lambda) = \frac{\|r_1\|_1 + \|r_2\|_1}{\lambda}$ *and*

$$C_1 = \frac{1}{(2\pi)^d} \int_{2\pi[-a/\lambda, a/\lambda]^d} f(\omega)^2 \big[|g(\omega)|^2 + g(\omega)\overline{g(-\omega)}\big] \, d\omega,$$

$$C_2 = \frac{1}{(2\pi)^d} \int_{2\pi[-a/\lambda, a/\lambda]^d} f(\omega)^2 \big[g(\omega)g(-\omega) + g(\omega)g(\omega)\big] \, d\omega.$$

Recall that $C_1 = C_1(a/\lambda)$ [*where* $C_1(\cdot)$ *is defined in* (3.4)].

In the following theorem, we derive bounds for the cumulants of $\widetilde{Q}_{a,\lambda}(g; r)$, which are subsequently used to show asymptotical normality of $\widetilde{Q}_{a,\lambda}(g; r)$.

THEOREM 4.4.   *Suppose Assumptions* 2.1, 2.2, 2.5 *and* 2.6(b) *hold. Then for all* $q \geq 3$ *and uniform in* $r_1, \ldots, r_q \in \mathbb{Z}^d$, *we have*

(4.5)
$$\text{cum}_q\big[\widetilde{Q}_{a,\lambda}(g, r_1), \ldots, \widetilde{Q}_{a,\lambda}(g, r_q)\big] = O\left(\frac{\log^{2d(q-2)}(a)}{\lambda^{d(q-1)}}\right)$$

if $\frac{\lambda^d}{n \log^{2d}(a)} \to 0$ *as* $n \to \infty$, $a \to \infty$ *and* $\lambda \to \infty$.

PROOF.   See [36], Section E.   □

From the above theorem, we see that if $\frac{\lambda^d}{n \log^{2d}(a)} \to 0$ and $\log^2(a)/\lambda^{1/2} \to 0$ as $\lambda \to \infty$, $n \to \infty$ and $a \to \infty$, then we have $\lambda^{dq/2} \, \text{cum}_q(\widetilde{Q}_{a,\lambda}(g, r_1), \ldots, \widetilde{Q}_{a,\lambda}(g, r_q)) \to 0$ for all $q \geq 3$. Using this result, we show asymptotic normality of $\widetilde{Q}_{a,\lambda}(g, r)$.

THEOREM 4.5.   *Suppose Assumptions* 2.1, 2.2, 2.5 *and* 2.6(b,c) *hold. Let* C_1 *and* C_2, *be defined as in Corollary* 4.1. *Under these conditions, we have*

$$\lambda^{d/2} \Delta^{-1/2} \begin{pmatrix} \Re\left(\widetilde{Q}_{a,\lambda}(g, r) - \langle \gamma, \gamma_{-r} \rangle I\left(g; \frac{a}{\lambda}\right)\right) \\ \Im\left(\widetilde{Q}_{a,\lambda}(g, r) - \langle \gamma, \gamma_{-r} \rangle I\left(g; \frac{a}{\lambda}\right)\right) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_2),$$

where

$$\Delta = \frac{1}{2} \begin{pmatrix} \Re(\Gamma_0 C_1 + \Gamma_{2r} C_2) & -\Im(\Gamma_{2r} C_2) \\ -\Im(\Gamma_{2r} C_2) & \Re(\Gamma_0 C_1 - \Gamma_{2r} C_2) \end{pmatrix}$$

with $\frac{\log^2(a)}{\lambda^{1/2}} \to 0$ *and* $\lambda^d/n \to 0$ *as* $\lambda \to \infty$, $n \to \infty$ *and* $a \to \infty$.

PROOF. See [36], Appendix E. □

It is likely that the above result also holds when the assumption of Gaussianity of the spatial random field is relaxed and replaced with the conditions stated in Theorem 4.6 (below) together with some mixing-type assumptions. We leave this for future work. However, in the following theorem, we obtain an expression for the variance of $\widetilde{Q}_{a,\lambda}(g; r)$ for non-Gaussian random fields.

THEOREM 4.6. *Let us suppose that* $\{Z(s); s \in \mathbb{R}^d\}$ *is a fourth-order station-ary spatial random field that satisfies Assumption* 2.1(i), 2.2, 2.5, 2.7 *and* 2.6(a,c) *or* 2.6(b,c) *are satisfied.*
If $\|r\|_1, \|r\|_2 \ll \lambda$, *then we have*

$$\lambda^d \operatorname{cov}[\widetilde{Q}_{a,\lambda}(g; r_1), \widetilde{Q}_{a,\lambda}(g; r_2)]$$
$$= \Gamma_{r_1-r_2}(C_1 + D_1) + O\left(\ell_{\lambda,a,n} + \frac{(a\lambda)^d}{n^2} + \varepsilon_{r_1,r_2}(\lambda)\right),$$

$$\lambda^d \operatorname{cov}[\widetilde{Q}_{a,\lambda}(g; r_1), \overline{\widetilde{Q}_{a,\lambda}(g; r_2)}]$$
$$= \Gamma_{r_1+r_2}(C_2 + D_2) + O\left(\ell_{\lambda,a,n} + \frac{(a\lambda)^d}{n^2} + \varepsilon_{r_1,r_2}(\lambda)\right),$$

where C_1 *and* C_2 *are defined as in Corollary* 4.1 *and*

$$D_1 = \frac{1}{(2\pi)^{2d}} \int_{2\pi[-a/\lambda, a/\lambda]^{2d}} g(\omega_1) \overline{g(\omega_2)} f_4(-\omega_1, -\omega_2, \omega_2) \, d\omega_1 \, d\omega_2,$$

$$D_2 = \frac{1}{(2\pi)^{2d}} \int_{2\pi[-a/\lambda, a/\lambda]^{2d}} g(\omega_1) g(\omega_2) f_4(-\omega_1, \omega_2, -\omega_2) \, d\omega_1 \, d\omega_2.$$

PROOF. See [36], Appendix C. □

We observe that to ensure the term $\frac{(a\lambda)^d}{n^2} \to 0$ we need to choose a such that $a^d = o(n^2/\lambda^d)$. In contrast for Gaussian random fields, $a = O(\lambda^k)$ for some $1 \le k < \infty$ was sufficient for obtaining an expression for the variance and asymptotic normality.

4.3. *Mixed domain verses pure increasing domain asymptotics.* The asymptotics in this paper are mainly done using mixed domain asymptotics, that is, as the domain $\lambda \to \infty$, the number of locations observed grows at a faster rate than λ, in other words $\lambda^d/n \to 0$ as $n \to \infty$. However, as rightly pointed out by a referee, for a given application it may be difficult to disambiguate Mixed Domain (MD) from the Pure Increasing Domain (PID) setup, where $\lambda^d/n \to c$ $(0 < c < \infty)$. We briefly discuss how the results change under PID asymptotics and the implications of this. We find that the results point to a rather intriguing difference for spatial processes that are Gaussian and non-Gaussian.

In the case that spatial process is Gaussian, using both MD and PID asymptotics we have $\lambda^d \operatorname{var}[\widetilde{Q}_{a,\lambda}(g; r)] = O(1)$ [see Theorem 4.3(i)]. Furthermore, an asymptotic expression for the variance is

$$\lambda^d \operatorname{var}[\widetilde{Q}_{a,\lambda}(g; r)] = \Gamma_0[C_1 + E_1] + O(\ell_{\lambda,a,n} + \varepsilon_{r,r}(\lambda)),$$

where $E_1 = O(\lambda^d/n)$ is a function of the spectral density; this term is not asymptotically negligible under PID asymptotics. From the above, we see that if we choose a such that $a = O(\lambda^k)$ for some $1 < k < \infty$ then similar results as those stated in Sections 4.1 and 4.2 hold under PID asymptotics. In the case that the process is non-Gaussian, using Theorem 4.6 we have

$$\lambda^d \operatorname{var}[\widetilde{Q}_{a,\lambda}(g; r)] = \Gamma_0[C_1 + D_1 + E_1 + F_1] + O\left(\ell_{\lambda,a,n} + \frac{(a\lambda)^d}{n^2} + \varepsilon_{r,r}(\lambda)\right),$$

where $F_1 = O(\lambda^d/n)$ is a function of the fourth-order spectral density function. However, there arises an additional term $O((a\lambda)^d/n^2)$. From the proof of Theorem 4.6, we see if $\frac{(a\lambda)^d}{n^2} \to \infty$ as $a, \lambda, n \to \infty$, then $\lambda^d \operatorname{var}[\widetilde{Q}_{a,\lambda}(g; r)]$ is not bounded. Thus, the number of frequencies, a, should be such that $(a\lambda)^d/n^2 \to 0$. In the case of MD asymptotics, we choose a such that $(a\lambda)^d/n^2 \to 0$ and $\log^3(a)/\lambda \to 0$. Under these two conditions, the frequency grid can be unbounded and grow at the rate a/λ as $\lambda \to \infty$. However, under PID asymptotics [where $\lambda = O(n^{1/d})$] in order to ensure that $(a\lambda)^d/n^2 = O(1)$ we require $a = O(n^{1/d}) = O(\lambda)$. This constrains the frequency grid to be bounded. To summarize, in the case that the spatial process is non-Gaussian and $n = O(\lambda^d)$ in order that $\operatorname{var}[\widetilde{Q}_{a,\lambda}(g; r)] \to 0$ as $\lambda \to \infty$, the frequency grid must be bounded or a coarser frequency grid $\omega_{\Omega,k}$ (where $\lambda > \Omega$) used (see Section 3).

4.4. *Fixed-domain asymptotics.* We now turn our attention to asymptotic sampling properties of $\widetilde{Q}_{a,\lambda}(g; 0)$ when the domain λ is kept fixed but the number of sampling locations $n \to \infty$. In order to simplify notation, we consider the case $d = 1$. We will assume the locations, $\{s_{n,j}\}$, lie close to a lattice and satisfy Assumption 2.4. It is clear that as $n \to \infty$ the Fourier transform $J_n(\omega_k)$ can be approximated by the Fourier transform over the continuum

$$\mathcal{J}_\lambda\left(\frac{k}{\lambda}\right) = \frac{1}{\lambda^{1/2}} \int_{-\lambda/2}^{\lambda/2} Z(s) \exp\left(\frac{2\pi i k s}{\lambda}\right) ds.$$

Thus, asymptotic expressions for the mean and variance of $\widetilde{Q}_{a,\lambda}(g; 0)$, where λ is fixed but $n \to \infty$ are in terms of the covariances of $\{\mathcal{J}_\lambda(\frac{k}{\lambda})\}_{k \in \mathbb{Z}}$. In Appendix G, [36] we show that

$$\text{var}\left[\mathcal{J}_\lambda\left(\frac{k}{\lambda}\right)\right] = A_\lambda\left(\frac{k}{\lambda}\right) \quad \text{and}$$

$$\text{cov}\left[\mathcal{J}_\lambda\left(\frac{k_1}{\lambda}\right), \mathcal{J}_\lambda\left(\frac{k_2}{\lambda}\right)\right] = \frac{(-1)^{k_1-k_2+1}}{\pi(k_1 - k_2)}\left[B_\lambda\left(\frac{k_1}{\lambda}\right) - B_\lambda\left(\frac{k_2}{\lambda}\right)\right],$$

where

(4.6)
$$A_\lambda\left(\frac{k}{\lambda}\right) = \int_{-\lambda}^{\lambda}\left(1 - \frac{|u|}{\lambda}\right)c(u)e^{2\pi i u/\lambda}\, du \quad \text{and}$$

$$B_\lambda\left(\frac{k}{\lambda}\right) = \int_0^\lambda c(u)\sin\left(\frac{2\pi ku}{\lambda}\right) du.$$

These expressions are used to prove the following result.

THEOREM 4.7. *Suppose Assumptions* 2.1, 2.4 *and* 2.6(e) *hold. Then keeping λ fixed but letting $n \to \infty$ we have*

$$(4.7) \quad \text{E}\left[\widetilde{Q}_{a,\lambda}(g; 0)\right] = \frac{1}{\lambda}\sum_{k=-a}^{a} g(\omega_k)A_\lambda\left(\frac{k}{\lambda}\right) + O\left(\frac{1}{n}\left[\sum_{k=-a}^{a}(|k| + 1)|g(\omega_k)|\right]\right)$$

and if $\sup_{\omega \in \mathbb{R}}|g(\omega)| < \infty$

(4.8)
$$\text{var}\left[\widetilde{Q}_{a,\lambda}(g; 0)\right] = \frac{1}{\lambda^2}\sum_{k_1,k_2=-\infty}^{\infty} g(\omega_k)\overline{g(\omega_{k_1})}B_\lambda(k_1, k_2)$$

$$+ O\left(\frac{a^4}{n^2} + \frac{a\log a}{n} + \frac{1}{a}\right),$$

where

$$B_\lambda(k_1, k_2) = \begin{cases} 2I(k=0)A_\lambda(0)^2 \\ \quad + I(k \neq 0)\left[A_\lambda\left(\frac{k}{\lambda}\right)^2 + \frac{1}{\pi^2 k^2}B_\lambda\left(\frac{k}{\lambda}\right)^2\right], & k_1 = k_2 (= k), \\ \frac{1}{\pi^2(k_1 - k_2)^2}\left[B_\lambda\left(\frac{k_1}{\lambda}\right) - B_\lambda\left(\frac{k_2}{\lambda}\right)\right]^2 \\ \quad + \frac{1}{\pi^2(k_1 + k_2)^2}\left[B_\lambda\left(\frac{k_1}{\lambda}\right) + B_\lambda\left(\frac{k_2}{\lambda}\right)\right]^2, & k_1 \neq k_2, \end{cases}$$

and $A_\lambda(\cdot)$ and $B_\lambda(\cdot)$ are defined in (4.6).

PROOF. See [36], Appendix G. □

From the above result, we see that if $\sup_{\omega \in \mathbb{R}} |g(\omega)| < \infty$ and the number of terms in the definition of $\widetilde{Q}_{a,\lambda}(g; 0)$, grows at a sufficiently slow rate as $n \to \infty$ then $\widetilde{Q}_{a,\lambda}(g; 0)$ is asymptotically an unbiased estimator of $\sum_{k \in \mathbb{Z}} g(\omega_k) A_\lambda(\frac{k}{\lambda})$ and the variance bounded [which fits with the conclusions of Theorem 4.3(ii)].

We now consider a special example of an estimator defined within the frequency domain and later compare it with the Gaussian maximum likelihood estimator for the same problem. More precisely, we consider the case that the covariance of a spatial Gaussian process is $\sigma^2 c(u)$ where $c(u)$ is known but σ^2 is unknown and our aim is to estimate σ^2. Let $A_\lambda(\frac{k}{\lambda}) = \int_{-\lambda}^{\lambda} c(u)(1 - |u|/\lambda) \exp(2\pi i k u/\lambda) \, du$. Since $E[|J_n(\omega_k)|^2] = \sigma^2 A_\lambda(\frac{k}{\lambda}) + O(n^{-1}(\lambda + |k|))$, it seems natural to use $\widehat{\sigma}^2$ as an estimator of σ^2, where

$$\widehat{\sigma}^2 = \frac{1}{a} \sum_{k=1}^{a} \frac{|J_n(\omega_k)|^2}{A_\lambda(\frac{k}{\lambda})}.$$

Note $\widehat{\sigma}^2$ corresponds to the Whittle likelihood estimator of σ^2 when $c(\cdot)$ is known. Using Theorem 4.7, equation (4.7) we have

$$E[\widehat{\sigma}^2] = \sigma^2 + O\left(\frac{1}{n}\left[\frac{1}{a}\sum_{k=1}^{a}(|k| + 1)|k|^2\right]\right).$$

Hence, if a is chosen such that the error above goes to zero as $n \to \infty$ then $\widehat{\sigma}^2$ is asymptotically an unbiased estimator of σ^2. Now we evaluate its variance. By using Corollary G.1, [36] it can be shown that

$$\text{(4.9)} \quad \begin{aligned} \text{var}[\widehat{\sigma}^2] &= \frac{1}{a^2} \sum_{k=1}^{a} \frac{B_\lambda(k, k)}{A_\lambda(\frac{k}{\lambda})^2} + \frac{2}{a^2} \sum_{k_1=1}^{a} \sum_{k_2=k_1+1}^{a} \frac{B_\lambda(k_1, k_2)}{A_\lambda(\frac{k_1}{\lambda}) A_\lambda(\frac{k_2}{\lambda})} \\ &\quad + O\left(\frac{a^6}{n^2} + \frac{a^3}{n}\right). \end{aligned}$$

But here we run into a problem. One would expect that if a is chosen such that $a^3/n \to 0$ as $a \to \infty$ and $n \to \infty$, then $\text{var}[\widehat{\sigma}^2] = O(a^{-1})$. This is true for the first term on the right-hand side of the above, *but* it is not necessarily true for the second term, which for most covariances will remain of order $O(1)$ for all a. The problem is that $A_\lambda(\frac{k}{\lambda})$ is a continuous function on the torus $[-1, 1]$, thus for large k $A_\lambda(\frac{k}{\lambda})$ decays at the rate k^{-2}. On the other hand, $B_\lambda(k_1/\lambda, k_2/\lambda)$ is a function of

$$B_\lambda\left(\frac{k_1}{\lambda}\right) - B_\lambda\left(\frac{k_2}{\lambda}\right) = \frac{1}{2}\int_0^\lambda c(u) \sin\left(\frac{k_1 - k_2}{\lambda}\pi u\right) \cos\left(\frac{k_1 + k_2}{\lambda}\pi u\right) du.$$

For most covariances $c(0) \neq c(\lambda)$, consequently, using integration by parts we see that $|B_\lambda(\frac{k_1}{\lambda}) - B_\lambda(\frac{k_2}{\lambda})| \sim |k_1 - k_2|^{-1}$ (a faster rate of convergence is not possible). Applying these bounds to (4.9) gives $\text{var}[\widehat{\sigma}^2] = O(1)$. Thus, even as $n \to \infty$ and $a \to \infty$, $\widehat{\sigma}^2$ is *not* a mean squared consistent estimator of σ^2. One exception is

when $c(u) = c(\lambda - u)$ for all $u \in [0, \lambda/2]$, in this case $B_\lambda(\frac{k}{\lambda}) = 0$ for all k and $\text{var}[\widehat{\sigma}^2] \approx a^{-1}$.

Therefore, in general, it seems that we cannot consistently estimate σ^2 using a Fourier domain approach. We conjecture that the only transformation of the data that will consistently estimate σ^2 is a transformation with the eigenfunctions associated with the covariance operator c. In contrast, [40] and [41] showed that if the maximum likelihood were used to estimate σ^2 in a Gaussian random field with covariance $\sigma^2 c(\cdot)$ where $c(\cdot)$ is a known Matern covariance function, then even within the fixed-domain framework σ^2 can be consistently estimated. This demonstrates that there exist situations where there are clear gains by working within the likelihood framework (if the correct distribution is specified). However, if the true covariance is $c(u) = c(u; \theta)$ and θ is also unknown, then even within the Gaussian likelihood framework one cannot consistently estimate σ^2 and θ.

5. A Studentized $\widetilde{Q}_{a,\lambda}(g; 0)$-statistic.

The expression for the variance $\widetilde{Q}_{a,\lambda}(g; 0)$ given in the examples above, is rather unwieldy and difficult to estimate directly. In this section we describe a simple method for estimating the variance of $\widetilde{Q}_{a,\lambda}(g; 0)$ under the assumption the locations are uniformly distributed. This estimator is used to obtain a simple Studentized statistic for $\widetilde{Q}_{a,\lambda}(g; 0)$. We assume in this section that $\widetilde{Q}_{a,\lambda}(g; 0)$ is a real random variable. Our approach is motivated by the method of orthogonal samples for time series proposed in [37], where the idea is to define a sample which by construction shares some of the properties as the estimator of interest. In this section we show that $\{\widetilde{Q}_{a,\lambda}(g; r); r \neq 0\}$ is an orthogonal sample associated with $\widetilde{Q}_{a,\lambda}(g; 0)$.

We will assume, for ease of presentation, that the spatial random field is Gaussian. Using Theorem 4.1 we have $\text{E}[\widetilde{Q}_{a,\lambda}(g; 0)] = I(g; \frac{a}{\lambda}) + o(1)$. Furthermore, noting that the Fourier coefficients of $h(\cdot)$ for uniformly sampled locations are $\gamma_0 = 1$ and $\gamma_j = 0$ if $j \neq 0$, using Theorem 4.3 we have

$$\lambda^d \text{var}[\widetilde{Q}_{a,\lambda}(g; 0)] = C_1 + O(\ell_{\lambda,a,n}).$$

In contrast, we observe that if no elements of the vector r are zero, then by Theorem 4.1 $\text{E}[\widetilde{Q}_{a,\lambda}(g; r)] = O(\prod_{i=1}^{d}[\log \lambda + \log |r_i|]/\lambda^d)$ (slightly slower rates are obtained when r contains zeros). In other words, $\widetilde{Q}_{a,\lambda}(g; r)$ is (asymptotically) estimating zero. On the other hand, using Theorem 4.3 we have

$$\lambda^d \text{cov}[\Re\widetilde{Q}_{a,\lambda}(g; r_1), \Re\widetilde{Q}_{a,\lambda}(g; r_2)]$$

$$= \begin{cases} \frac{1}{2}C_1 + O\left(\ell_{\lambda,a,n} + \dfrac{\|r\|_1}{\lambda}\right), & r_1 = r_2 (= r), \\ O(\ell_{\lambda,a,n}), & r_1 \neq r_2, r_1 \neq -r_2. \end{cases}$$

A similar result holds for $\{\Im\widetilde{Q}_{a,\lambda}(g; r)\}$, furthermore we have $\lambda^d \text{cov}[\Re\widetilde{Q}_{a,\lambda}(g; r_1), \Im\widetilde{Q}_{a,\lambda}(g; r_2)] = O(\ell_{\lambda,a,n})$.

In summary, if $\|\boldsymbol{r}\|_1$ is not too large, then $\{\Re\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r}), \Im\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})\}$ are "near uncorrelated" random variables whose variance is approximately the same as $\widetilde{Q}_{a,\lambda}(g;0)/\sqrt{2}$. This suggests we use $\{\Re\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r}), \Im\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r}); \boldsymbol{r} \in \mathcal{S}\}$ to estimate $\mathrm{var}[\widetilde{Q}_{a,\lambda}(g;0)]$, where the set $\mathcal{S}$ is defined as

(5.1) $\mathcal{S} = \{\boldsymbol{r}; \|\boldsymbol{r}\|_1 \leq M, r_1 \neq r_2 \text{ and all elements of } \boldsymbol{r} \text{ are nonzero}\}.$

This leads to the following estimator:

(5.2)
$$\widetilde{V}_{\mathcal{S}} = \frac{\lambda^d}{2|\mathcal{S}|} \sum_{\boldsymbol{r}\in\mathcal{S}} \left(2|\Re\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})|^2 + 2|\Im\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})|^2\right)$$
$$= \frac{\lambda^d}{|\mathcal{S}|} \sum_{\boldsymbol{r}\in\mathcal{S}} |\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})|^2,$$

where $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$. Note that we specifically select the set $\mathcal{S}$ such that no element $\boldsymbol{r}$ contains zero, this is to ensure that $\mathrm{E}[\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})]$ is small and does not induce a large bias in $\widetilde{V}_{\mathcal{S}}$

In the following theorem, we obtain a mean squared bound for $\widetilde{V}_{\mathcal{S}}$.

THEOREM 5.1.   *Let $\widetilde{V}_{\mathcal{S}}$ be defined as in* (5.2), *where $\mathcal{S}$ is defined in* (5.1). *Suppose Assumptions* 2.1, 2.3 *and* 2.6(a,b,c) *hold and either Assumption* 2.5(i) *or* (ii) *holds. Then we have*

$$\mathrm{E}\left(\widetilde{V}_{\mathcal{S}} - \lambda^d\,\mathrm{var}[\widetilde{Q}_{a,\lambda}(g;0)]\right)^2 = O\left(|\mathcal{S}|^{-1} + |M|\lambda^{-1} + \ell_{\lambda,a,n} + \lambda^{-d}\log^{4d}(a)\right)$$

as $\lambda \to \infty$, $a \to \infty$ and $n \to \infty$ [where $\ell_{a,\lambda,n}$ is defined in (4.3)*].*

PROOF.   See [36], Appendix I.   □

Thus, it follows from the above result that if the set $\mathcal{S}$ grows at a rate such that $|M|\lambda^{-1} \to 0$ as $\lambda \to \infty$, then $\widetilde{V}_{\mathcal{S}}$ is a mean square consistent estimator of $\lambda^d\,\mathrm{var}[\widetilde{Q}_{a,\lambda}(g;0)]$. We use this result to define an asymptotically pivotal statistic. Let

$$T_{\mathcal{S}} = \frac{\lambda^{d/2}[\widetilde{Q}_{a,\lambda}(g;0) - I(g;\frac{a}{\lambda})]}{\sqrt{\widetilde{V}_{\mathcal{S}}}}.$$

By using Theorem E.1 in [36], we can immediately show that for fixed $\mathcal{S}$, $T_{\mathcal{S}} \xrightarrow{\mathcal{D}} t_{2|\mathcal{S}|}$ as $\lambda \to \infty$. Therefore, $T_{\mathcal{S}}$ is asymptotically pivotal and can be used to construct confidence intervals and test hypothesis about the parameter $I(g;\frac{a}{\lambda})$.

We note that the same approach and Studentization can be used in the case that the random field is non-Gaussian. However, it is unclear how to relax the assumption that the locations are uniformly distributed. This is because in the case of a nonuniform design $\mathrm{E}[\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r})]$ ($\boldsymbol{r} \neq 0$) will not, necessarily, be estimating zero.

Acknowledgments. The author gratefully acknowledges Gregory Berkolaiko and Mikyoung Jun for many very useful discussions. Furthermore, the author gratefully acknowledges the Editor, Associate Editor and three anonymous referees who made substantial improvements and tirelessly corrected every part of the manuscript.

SUPPLEMENTARY MATERIAL

Supplement to "Statistical inference for spatial statistics defined in the Fourier domain" (DOI: 10.1214/17-AOS1556SUPP; .pdf). The supplement contains the proofs for all the results of the main article and some related results.

REFERENCES

[1] BANDYOPADHYAY, S. and LAHIRI, S. N. (2009). Asymptotic properties of discrete Fourier transforms for spatial data. *Sankhyā* **71** 221–259. MR2639292

[2] BANDYOPADHYAY, S., LAHIRI, S. N. and NORDMAN, D. J. (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *Ann. Statist.* **43** 519–545. MR3316189

[3] BANDYOPADHYAY, S. and SUBBA RAO, S. (2017). A test for stationarity for irregularly spaced spatial data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 95–123. MR3597966

[4] BEUTLER, F. J. (1970). Alias-free randomly timed sampling of stochastic processes. *IEEE Trans. Inform. Theory* **16** 147–152. MR0273761

[5] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhya, Ser. A* **50** 381–393. MR1065550

[6] BRILLINGER, D. (1969). The calculation of cumulants via conditioning. *Ann. Inst. Statist. Math.* **21** 215–218.

[7] BRILLINGER, D. R. (1981). *Time Series: Data Analysis and Theory*, 2nd ed. Holden-Day, Inc., Oakland, CA. MR0595684

[8] CAN, S. U., MIKOSCH, T. and SAMORODNITSKY, G. (2010). Weak convergence of the function-indexed integrated periodogram for infinite variance processes. *Bernoulli* **16** 995–1015. MR2759166

[9] CHEN, W. W., HURVICH, C. M. and LU, Y. (2006). On the correlation matrix of the discrete Fourier transform and the fast solution of large Toeplitz systems for long-memory time series. *J. Amer. Statist. Assoc.* **101** 812–822. MR2281249

[10] CRESSIE, N. and HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94** 1330–1340. MR1731494

[11] CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York. Revised reprint of the 1991 edition. MR1239641

[12] DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766. MR1026311

[13] DAHLHAUS, R. and JANAS, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *Ann. Statist.* **24** 1934–1963. MR1421155

[14] DEO, R. S. and CHEN, W. W. (2000). On the integral of the squared periodogram. *Stochastic Process. Appl.* **85** 159–176. MR1730613

[15] DUNSMUIR, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Ann. Statist.* **7** 490–506. MR0527485

[16] FOX, R. and TAQQU, M. S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields* **74** 213–240. MR0871252

[17] FUENTES, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Amer. Statist. Assoc.* **102** 321–331. MR2345545

[18] GINÉ, E. and NICKL, R. (2008). A simple adaptive estimator of the integrated square of a density. *Bernoulli* **14** 47–61. MR2401653

[19] GIRAITIS, L. and SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields* **86** 87–104. MR1061950

[20] HALL, P., FISHER, N. I. and HOFFMANN, B. (1994). On the nonparametric estimation of covariance functions. *Ann. Statist.* **22** 2115–2134. MR1329185

[21] HALL, P. and PATIL, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab. Theory Related Fields* **99** 399–424. MR1283119

[22] HANNAN, E. J. (1971). Non-linear time series regression. *J. Appl. Probab.* **8** 767–780. MR0295498

[23] KAWATA, T. (1959). Some convergence theorems for stationary stochastic processes. *Ann. Math. Stat.* **30** 1192–1214. MR0109365

[24] LAHIRI, S. N. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhyā* **65** 356–388. MR2028905

[25] LAURENT, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981

[26] MASRY, E. (1978). Poisson sampling and spectral estimation of continuous-time processes. *IEEE Trans. Inform. Theory* **24** 173–183. MR0483238

[27] MATSUDA, Y. and YAJIMA, Y. (2009). Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 191–217. MR2655530

[28] NEIBUHR, T. and KREISS, J.-P. (2014). Asymptotics for autocovariances and integrated periodograms for linear processes observed at lower frequencies. *Int. Stat. Rev.* **82** 123–140. MR3200539

[29] PARZEN, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Stat.* **28** 329–348. MR0088833

[30] PECCATI, G. and TAQQU, M. S. (2011). *Wiener Chaos*: *Moments*, *Cumulants and Diagrams*: *A Survey with Computer Implementation*. *Bocconi & Springer Series* **1**. Springer, Milan; Bocconi Univ. Press, Milan. MR2791919

[31] RICE, J. (1979). On the estimation of the parameters of a power spectrum. *J. Multivariate Anal.* **9** 378–392. MR0548788

[32] SHAPIRO, H. S. and SILVERMAN, R. A. (1960). Alias-free sampling of random noise. *J. Soc. Ind. Appl. Math.* **8** 225–248. MR0121948

[33] STEIN, M. L. (1995). Fixed-domain asymptotics for spatial periodograms. *J. Amer. Statist. Assoc.* **90** 1277–1288. MR1379470

[34] STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. Springer, New York. MR1697409

[35] STEIN, M. L., CHI, Z. and WELTY, L. J. (2004). Approximating likelihoods for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 275–296. MR2062376

[36] SUBBA RAO, S. (2018). Supplement to "Statistical inference for spatial statistics defined in the Fourier domain." DOI:10.1214/17-AOS1556SUPP.

[37] SUBBA RAO, S. (2017). Orthogonal samples for estimators in time series. Under revision.

[38] VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. R. Stat. Soc., B* **50** 297–312. MR0964183

[39] WALKER, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Aust. Math. Soc.* **4** 363–384. MR0171345

[40] ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303

[41] ZHANG, H. and ZIMMERMAN, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. MR2234195

DEPARTMENT OF STATISTICS
TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843
USA
E-MAIL: suhasini@stat.tamu.edu