The Debiased Spatial Whittle Likelihood

Arthur P. Guillaumin[1], Adam M. Sykulski[2], Sofia C. Olhede[3,1]
and Frederik J. Simons[4]

[1] *University College London, UK*
[2] *Lancaster University, UK*
[3] *École polytechnique fédérale de Lausanne, Swizterland*
[4] *Princeton University, USA*

Abstract

We provide a computationally and statistically efficient method for estimating the parameters of a stochastic Gaussian model observed on a regular spatial grid in any number of dimensions. Our proposed method, which we call the debiased spatial Whittle likelihood, makes important corrections to the well-known Whittle likelihood to account for large sources of bias caused by boundary effects and aliasing. We generalise the approach to flexibly allow for significant volumes of missing data, for the usage of irregular sampling schemes including those with lower-dimensional substructure, and for irregular sampling boundaries. We build a theoretical framework under relatively weak assumptions which ensures consistency and asymptotic normality in numerous practical settings. We provide detailed implementation guidelines which ensure the estimation procedure can still be conducted in $\mathcal{O}(n \log n)$ operations, where n is the number of points of the encapsulating rectangular grid, thus keeping the computational scalability of Fourier and Whittle-based methods for large data sets. We validate our procedure over a range of simulated and real world settings, and compare with state-of-the-art alternatives, demonstrating the enduring significant practical appeal of Fourier-based methods, provided they are corrected by the constructive procedures developed in this paper.

Keywords: Random fields; Missing data; Irregular boundaries; Aliasing; Whittle likelihood

1 Introduction

Among the main challenges of modern data analysis is making sense of large volumes of spatial and spatiotemporal data. The state-of-the-art parameter estimation methods currently in use are based on various different likelihood approximation methods designed to combine statistical and computational efficiency. Such methods are primarily reliant on spatial/pixel

models (Anitescu et al., 2017; Guinness and Fuentes, 2017; Katzfuss, 2017; Stroud et al., 2017), spectral/Fourier understanding (Kaufman et al., 2008; Matsuda and Yajima, 2009; Shaby and Ruppert, 2012; Guinness, 2019), or other methods of likelihood approximation (Stein et al., 2004; Banerjee et al., 2008; Lee and Mitchell, 2013; Sang and Huang, 2012). Fourier-based methods, typically based on the Whittle likelihood, are fast and can scale well to massive data sets. Fourier-based methods, on the other hand, are known to engender large sources of bias, particularly in dimensions greater than one (Dahlhaus and Künsch, 1987), or in the presence of missing data or under irregular sampling (Fuentes, 2007; Matsuda and Yajima, 2009). In this paper we propose a novel methodology that simultaneously addresses these challenges for spatial data observed on a regular grid, which may have missing data and irregular sampling boundaries, in any number of dimensions.

The bias which we remove is due to finite-domain effects, and to the multidimensional boundary. Much of the literature on Whittle estimation has focused on modifications to the periodogram to reduce bias, such as tapering (Dahlhaus and Künsch, 1987), edge-effect estimation (Robinson and Sanz, 2006), or accounting for non-standard sampling scenarios (Fuentes, 2007; Matsuda and Yajima, 2009; Rao, 2018). The solution we propose is simple yet effective: determine the true expectation of the periodogram, and construct a quasi-likelihood using this quantity rather than the true spectrum—further developing and generalizing a procedure recently proposed by Sykulski et al. (2019) for one-dimensional completely observed time series. We shall show that the debiased Spatial Whittle likelihood almost completely removes estimation bias in spatial inference, even in the presence of significant amounts of missing data, while leaving estimation variance essentially unaffected. We also establish a convergence rate under very general sampling and model assumptions.

The notion of debiasing Whittle estimates using the expected periodogram has been previously investigated in various more restrictive frameworks by Fernández-Casal and Crujeiras (2010), Simons and Olhede (2013), and Deb et al. (2017). This article however is the first to formalize the estimation procedure by providing theoretical guarantees that apply to any number of dimensions, allow for missing data, and account for irregular sampling boundaries. To achieve this we introduce the concept of *significant correlation contribution*, which provides weak conditions on sampling regimes for stochastic process models that allow for asymptotically efficient parameter estimation. This paper is also the first to provide fast $n \log n$ computational implementation, including for missing data and higher dimensions.

This article first provides the choice of notation and assumptions in Section 2. We propose our spatial quasi-likelihood in Section 3. Then in Section 4 we introduce significant correlation contribution, which provides conditions guaranteeing consistent estimation under a wide range of sampling schemes. Section 5 develops our theoretical results. Section 6 shows the improved performance on simulated data and actual data of Venus' topography. Finally we conclude with discussion in Section 7.

2 Notation and assumptions

Consider a finite-variance and zero-mean Gaussian random field $X(\mathbf{s})$, for $\mathbf{s} \in \mathbb{R}^d$, where $d \geq 1$ is a positive integer. Under the assumption of homogeneity, we denote the covariance function of $X(\mathbf{s})$ by $c_X(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^d$, and assume the existence of a positive piecewise continuous Riemann-integrable spectral density function $f_X(\boldsymbol{\omega})$, such that $\forall \mathbf{u}, \mathbf{s} \in \mathbb{R}^d$,

$$c_X(\mathbf{u}) = \mathrm{E}\left\{X(\mathbf{s})X(\mathbf{s}+\mathbf{u})\right\} = \int_{\mathbb{R}^d} f_X(\boldsymbol{\omega}) \exp(i\boldsymbol{\omega} \cdot \mathbf{u}) \, d\boldsymbol{\omega}, \tag{1}$$

and $f_X(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} c_X(\mathbf{u}) \exp(-i\boldsymbol{\omega} \cdot \mathbf{u}) \, d\mathbf{u}$. We shall assume the spectral density belongs to a parametric family indexed by the parameter $\boldsymbol{\theta} \in \Theta$, or $f_X(\boldsymbol{\omega}) = f(\boldsymbol{\omega}; \boldsymbol{\theta}_0)$, denoting the true parameter value by $\boldsymbol{\theta}_0 \in \Theta$. The random field $X(\mathbf{s})$ is taken to be Gaussian and homogeneous but not necessarily isometric. We denote $\mathbf{n} = (n_1, \ldots, n_d) \in (\mathbb{N}^+)^d$, with $\mathbb{N}^+$ the set of positive integers, the dimensions of an orthogonal regular and rectangular *bounding grid*, defined by

$$\mathcal{J}_\mathbf{n} = \left\{\boldsymbol{\delta} \circ [x_1, \ldots, x_d]^T : (x_1, \ldots, x_d) \in \mathbb{N}^d, \ 0 \leq x_i \leq n_i - 1, i = 1, \ldots, d\right\}, \tag{2}$$

and denote by $|\mathbf{n}| = \prod_{i=1}^d n_i$ the total number of points of this grid. We denote by $X_\mathbf{s}$, $\mathbf{s} \in \mathcal{J}_\mathbf{n}$ the values of the process on the grid. In (2), the quantity $\boldsymbol{\delta} \in (\mathbb{R}^+)^d$ indicates the regular spacing along each axis, with $\mathbb{R}^+$ the set of positive real numbers, and $\circ$ denotes the pointwise Hadamard product between two vectors. We always take $\boldsymbol{\delta} = [1, \ldots, 1]^T$ for simplicity, yet without loss of generality. We write $f_{X,\boldsymbol{\delta}}(\boldsymbol{\omega})$ for the spectral density of the sampled process, the *aliased* spectral density, defined by

$$f_{X,\boldsymbol{\delta}}(\boldsymbol{\omega}) = \sum_{\mathbf{u} \in \mathbb{Z}^d} f_X(\boldsymbol{\omega} + 2\pi\mathbf{u}), \qquad \boldsymbol{\omega} \in \mathbb{R}^d, \tag{3}$$

which is a Fourier pair with $c_X(\mathbf{u}) = \int_{\mathcal{T}^d} f_{X,\delta}(\boldsymbol{\omega}) \exp(i\boldsymbol{\omega} \cdot \mathbf{u}) \, d\boldsymbol{\omega}$, $\forall \mathbf{u} \in \mathbb{Z}^d$, and $\mathcal{T} = [0, 2\pi)$, with $\mathbb{Z}$ the set of natural integers.

To account for irregular domain shapes and missing data, we define a deterministic modulation value $g_\mathbf{s}$ at each location of the grid $\mathcal{J}_\mathbf{n}$. If a point on the regular grid is missing then $g_\mathbf{s} = 0$, otherwise $g_\mathbf{s} = 1$. By convention, $g_\mathbf{s}$ is extended to the whole set $\mathbb{Z}^d$, defining $g_\mathbf{s} = 0$ if $\mathbf{s} \notin \mathcal{J}_\mathbf{n}$. Using this notation, the periodogram of the observed data takes the form

$$I_\mathbf{n}(\boldsymbol{\omega}) = \frac{(2\pi)^{-d}}{\sum_{\mathbf{s} \in \mathcal{J}_\mathbf{n}} g_\mathbf{s}^2} \left| \sum_{\mathbf{s} \in \mathcal{J}_\mathbf{n}} g_\mathbf{s} X_\mathbf{s} \exp(-i\boldsymbol{\omega} \cdot \mathbf{s}) \right|^2, \qquad \boldsymbol{\omega} \in \mathbb{R}^d, \tag{4}$$

where normalizing by $\sum_{\mathbf{s} \in \mathcal{J}_\mathbf{n}} g_\mathbf{s}^2$ rescales the periodogram for missing data, as performed in Fuentes (2007). Note that, despite this similarity, our approach is fundamentally different to that of Fuentes (2007), where this extended definition of the periodogram is used in the Whittle procedure to address missing data. While this rescaling is central to the method proposed by Fuentes, it is merely a convention in our case. In practice, this rescaling is not actually required in our implementation.

Evaluating the periodogram on the multidimensional Fourier grid

$$\Omega_{\mathbf{n}} = \prod_{i=1}^{d} \left\{ 2\pi k n_i^{-1} : k = 0, \ldots, n_i \right\}$$

associated with the spatial grid $\mathcal{J}_{\mathbf{n}}$ requires $\mathcal{O}(|\mathbf{n}| \log |\mathbf{n}|)$ elementary operations using the Fast Fourier Transform (FFT). If a taper is used in the spectral estimate, then the values of the taper are directly incorporated into $g_{\mathbf{s}}$, such that $g_{\mathbf{s}}$ is proportional to the taper at locations where data are observed (and still set to zero otherwise). We shall assume that $g_{\mathbf{s}}$ takes values in the interval $[0, 1]$, however this condition could be relaxed to assuming an upper-bound for the absolute value.

3 Methodology

In Section 3.1 we introduce our new spatial frequency-domain quasi-likelihood estimator. In Section 3.2 we present an algorithm for the computation of the quasi-likelihood that only requires FFTs, even in the scenario of missing data and general boundaries. Thus our estimation method uniquely retains the $\mathcal{O}(|\mathbf{n}| \log |\mathbf{n}|)$ computational cost of frequency-domain approaches for regular grids. Finally, in Section 4 we introduce the notion of *significant correlation contribution* for random fields, a set of sufficient conditions that ensures that our estimate is consistent under growing-domain asymptotics, for a wide range of sampling schemes and model families.

3.1 Estimation procedure

While exact likelihood has optimal statistical properties in the framework of an increasing domain (Mardia and Marshall, 1984), it is computationally inadequate for large data sets of spatial observations due to the determinant calculation and linear system that needs to be solved. A common approach is to trade off computational cost with statistical efficiency by using approximations of the likelihood function (Fuentes, 2007; Varin et al., 2011; Guinness and Fuentes, 2017). Such functions are commonly called quasi-likelihood methods. Our proposed estimation method is based on the following quasi-likelihood,

$$\ell(\boldsymbol{\theta}) = |\mathbf{n}|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}}} \left\{ \log \overline{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\theta}) + \frac{I_{\mathbf{n}}(\boldsymbol{\omega})}{\overline{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\theta})} \right\}, \quad \forall \boldsymbol{\omega} \in \mathcal{T}^d, \tag{5}$$

where for all $\boldsymbol{\theta} \in \Theta$,

$$\overline{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}\{I_{\mathbf{n}}(\boldsymbol{\omega})\}, \tag{6}$$

is defined as the expected periodogram given the modulation values $g_{\mathbf{s}}$, under the mean-zero Gaussian distribution of $X_{\mathbf{s}}$ specified by the parameter vector $\boldsymbol{\theta}$—see also Fernández-Casal and Crujeiras (2010). Replacing $\overline{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\theta})$ with $f_X(\boldsymbol{\omega})$ in (5) yields the discretized form of the standard Whittle likelihood. Note however that unlike the spectral density, the

expected periodogram directly accounts for the sampling, as it depends on the dimensions of the lattice $\mathbf{n}$ as well as the modulation values $g_{\mathbf{s}}$ accounting for missing points. We minimize (5) over Θ to obtain our estimate,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta}\{\ell(\boldsymbol{\theta})\}. \tag{7}$$

By minimizing (5), we find the maximum-likelihood estimate of the data under the following parametric model,

$$I_{\mathbf{n}}(\boldsymbol{\omega}) \overset{i.i.d.}{\sim} \mathrm{Exp}\left\{\overline{I}_{\mathbf{n}}(\boldsymbol{\omega};\boldsymbol{\theta})^{-1}\right\}, \qquad \boldsymbol{\omega} \in \Omega_{\mathbf{n}}, \tag{8}$$

where $\mathrm{Exp}(\lambda)$ stands for the exponential distribution with parameter λ. Hence the quantity given in (5) can be seen as a composite likelihood (Varin et al., 2011; Bevilacqua and Gaetan, 2015). We also observe that $\Delta_{\boldsymbol{\theta}}\ell_M(\boldsymbol{\theta}_0) = \mathbf{0}$ such that our method fits within the general theory of estimating equations (Heyde, 1997; Jesus and Chandler, 2017).

3.2 Computation of the expected periodogram

In this section we show how the expected periodogram in (5) can be computed using FFTs such that our quasi-likelihood remains an $\mathcal{O}(|\mathbf{n}|\log|\mathbf{n}|)$ procedure, for any dimension d and independently of the missing data patterns. Direct calculations show that the expected periodogram is the convolution of the spectral density of the process with the multi-dimensional kernel $\mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega})$,

$$\overline{I}_{\mathbf{n}}(\boldsymbol{\omega};\boldsymbol{\theta}) = \{f_X(\,\cdot\,;\boldsymbol{\theta}) * \mathcal{F}_{\mathbf{n}}(\cdot)\}(\boldsymbol{\omega}) = \int_{\mathcal{T}^d} f_{X,\boldsymbol{\delta}}(\boldsymbol{\omega} - \boldsymbol{\omega}';\boldsymbol{\theta})\mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega}')\,d\boldsymbol{\omega}',$$

where

$$\mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega}) = \frac{(2\pi)^{-d}}{\sum g_{\mathbf{s}}^2}\left|\sum_{\mathbf{s} \in \mathcal{J}_{\mathbf{n}}} g_{\mathbf{s}}\exp(i\boldsymbol{\omega} \cdot \mathbf{s})\right|^2, \qquad \boldsymbol{\omega} \in \mathbb{R}^d. \tag{9}$$

When $g_{\mathbf{s}} = 1 \; \forall \mathbf{s} \in \mathcal{J}_{\mathbf{n}}$, $\mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega})$ is simply the multi-dimensional rectangular Féjer kernel, i.e. a separable product of one-dimensional Féjer kernels. For this reason we call $\mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega})$ a *modified* Féjer kernel. We now provide two lemmata stating that the expected periodogram can be computed via FFTs for any value of the modulation $g_{\mathbf{s}}$ on the grid $\mathcal{J}_{\mathbf{n}}$.

Lemma 1 (Expected periodogram as a Fourier series)**.** *The expected periodogram can be written as the following Fourier series,*

$$\overline{I}_{\mathbf{n}}(\boldsymbol{\omega};\boldsymbol{\theta}) = (2\pi)^{-d}\sum_{\mathbf{u} \in \mathbb{Z}^d}\overline{c}_{\mathbf{n}}(\mathbf{u};\boldsymbol{\theta})\exp(-i\boldsymbol{\omega} \cdot \mathbf{u}), \quad \forall \boldsymbol{\omega} \in \mathcal{T}^d, \forall \boldsymbol{\theta} \in \Theta, \tag{10}$$

where $\overline{c}_{\mathbf{n}}(\mathbf{u};\boldsymbol{\theta})$ is defined by,

$$\overline{c}_{\mathbf{n}}(\mathbf{u};\boldsymbol{\theta}) = c_{g,\mathbf{n}}(\mathbf{u})c_X(\mathbf{u};\boldsymbol{\theta}), \qquad \mathbf{u} \in \mathbb{Z}^d, \text{with,} \tag{11}$$

$$c_{g,\mathbf{n}}(\mathbf{u}) = \frac{\sum_{\mathbf{s} \in \mathcal{J}_{\mathbf{n}}} g_{\mathbf{s}}g_{\mathbf{s}+\mathbf{u}}}{\sum_{\mathbf{s} \in \mathcal{J}_{\mathbf{n}}} g_{\mathbf{s}}^2}, \qquad \mathbf{u} \in \mathbb{Z}^d. \tag{12}$$

5

Proof. This is obtained by direct calculation on taking the expectation of the periodogram as defined in (4). □

In practice we can evaluate the expected periodogram at the set of Fourier frequencies through a multidimensional FFT, as detailed in the following lemma.

Lemma 2 (Computation of the expected periodogram via FFT). *The expected periodogram takes the form*

$$\overline{I}_{\mathbf{n}}(\boldsymbol{\omega}_k; \boldsymbol{\theta}) = (2\pi)^{-d} \sum_{u_1=0}^{n_1-1} \cdots \sum_{u_d=0}^{n_d-1} \sum_{\mathbf{q}} \overline{c}_{\mathbf{n}} (\mathbf{u} - \mathbf{q} \circ \mathbf{n}; \boldsymbol{\theta}) \exp(-i\boldsymbol{\omega}_k \cdot \mathbf{u}), \qquad (13)$$

where the sum over $\mathbf{q}$ *ranges over all vectors of size* d *with elements in the set* $\{0, 1\}$ *(hence,* 2^d *of them), and where* $\circ$ *denotes the Hadamard product. Thus the expected periodogram can be computed via an aggregation of* 2^d *separate FFTs.*

Proof. Please see the Supplementary material. □

In dimension $d = 2$, q goes over $\left\{ [0 \ 0]^T, [1 \ 1]^T, [0 \ 1]^T, [1 \ 1]^T \right\}$, and the above formula therefore takes the form

$$\overline{I}_{\mathbf{n}}(\boldsymbol{\omega}_k; \boldsymbol{\theta}) = (2\pi)^{-d} \sum_{u_1=0}^{n_1-1} \cdots \sum_{u_d=0}^{n_d-1} \{ \overline{c}_{\mathbf{n}} ([u_1 \ u_2]; \boldsymbol{\theta}) + \overline{c}_{\mathbf{n}} ([u_1 - n_1 \ u_2 - n_2]; \boldsymbol{\theta}) \qquad (14)$$
$$+ \ \overline{c}_{\mathbf{n}} ([u_1 \ u_2 - n_2]; \boldsymbol{\theta}) + \overline{c}_{\mathbf{n}} ([u_1 - n_1 \ u_2]; \boldsymbol{\theta}) \} \exp(-i\boldsymbol{\omega}_k \cdot \mathbf{u}).$$

We remind the reader that $g_{\mathbf{s}}$ is defined to be zero outside $\mathcal{J}_{\mathbf{n}}$. Hence $c_{g,\mathbf{n}}(\mathbf{u})$ is the ratio of the number of pairs of observations *separated* by the vector $\mathbf{u}$ over the total number of points of the rectangular grid $\mathcal{J}_{\mathbf{n}}$. In the special case of complete observations on the rectangular grid (12) simplifies to

$$c_{g,\mathbf{n}}(\mathbf{u}) = \begin{cases} |\mathbf{n}|^{-1} \prod_{i=1}^{d} (n_i - |u_i|) = \prod_{i=1}^{d} \left(1 - \frac{|u_i|}{n_i}\right) & \text{if } |u_i| \leq n_i - 1, i = 1, \ldots, d, \\ 0 & \text{otherwise.} \end{cases} \qquad (15)$$

In the general case, $c_{g,\mathbf{n}}(\mathbf{u})$ is precomputed for all relevant values of $\mathbf{u}$ via an FFT independently of the parameter value $\boldsymbol{\theta}$, such that our method can be applied to scenarios of missing data without loss of computational efficiency. Similarly, we can combine our debiasing procedure with tapering by using a tapered spectral estimate for $I_{\mathbf{n}}(\boldsymbol{\omega})$ in (5) with adjusted values for $g_{\mathbf{s}}$ (as discussed at the end of Section 2). The expected periodogram, $\overline{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\theta})$, is then computed by using these values of $g_{\mathbf{s}}$ in the formulation of $c_{g,\mathbf{n}}(\mathbf{u})$ in (12). Combining debiasing and tapering therefore remains an $\mathcal{O}(|\mathbf{n}| \log |\mathbf{n}|)$ procedure.

4 Properties of sampling patterns

To account for missing observations on the rectangular grid $\mathcal{J}_{\mathbf{n}}$, we have replaced missing values with zeros via the modulation function $g_{\mathbf{s}}$. Depending on $g_{\mathbf{s}}$ this may result in losing identifiability of the parameter vector from the second-order moment quantities available from the data. More generally, we wish to understand how the sampling pattern affects the consistency of our estimation procedure. To this end, we define the notion of significant correlation contribution for spatial random fields, which allows us to determine whether our sampling pattern samples enough *spatial lags* where information about the model lies. This generalizes ideas from modulated time series (Guillaumin et al., 2017). Following two simple lemmata on some properties of $c_{g,\mathbf{n}}(\mathbf{u})$, we provide the formal definition of Significant Correlation Contribution (SCC), and follow with some general cases and an example with an isometric model family to provide more intuition and demonstrate the generality of our framework.

4.1 Basic properties of $c_g(\mathbf{u})$ and $\mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega})$

We state three basic properties of the introduced quantity $c_{g,\mathbf{n}}(\mathbf{u})$, both in order to provide more intuition, but also for further use in this paper.

Lemma 3. *The kernel $\mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega})$ defined in (9) and $c_{g,\mathbf{n}}(\mathbf{u})$ form a Fourier pair.*

Lemma 4. *We have*
$$0 \leq c_{g,\mathbf{n}}(\mathbf{u}) \leq 1, \quad \forall \mathbf{u} \in \mathbb{Z}. \tag{16}$$

Proof. The left side of the inequality is obvious as by assumption $g_{\mathbf{s}} \geq 0$. The right side is obtained by direct application of the Cauchy-Schwartz inequality. $\square$

Lemma 5. *We have*
$$\sum_{\mathbf{u} \in \mathbb{Z}} c_{g,\mathbf{n}}(\mathbf{u}) \geq \sum_{\mathbf{s} \in \mathbb{Z}} g_{\mathbf{s}}. \tag{17}$$

Proof. Please see the Supplementary material. $\square$

4.2 Definitions

Our concept of Significant Correlation Contribution (SCC), which we provide below, is defined in asymptotic terms, since we shall later on make use of it to establish consistency of our estimator. More specifically, we consider a sequence of grids, indexed by $k \in \mathbb{N}$, which goes to infinity, rather than a single grid.

Definition 1 (Significant Correlation Contribution (SCC)). *A sequence of observed grids $(\mathcal{J}_{\mathbf{n}_k}, g_k)_{k \in \mathbb{N}}$ leads to significant correlation contribution for the model family $\{f(\,\cdot\,; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ if it satisfies both*

$$\begin{cases} \sum_{\mathbf{u} \in \mathbb{Z}} c_{g,\mathbf{n}_k}(\mathbf{u}) c_X^2(\mathbf{u}) \underset{k \to \infty}{=} o\left(\sum g_{\mathbf{s}}^2\right), \\ \underline{\lim}_{k \to \infty} S_k(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > 0, \quad \forall \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \in \Theta, \end{cases} \tag{18}$$

where $\underline{\lim}_{k\to\infty}$ denotes the limit inferior and where we have defined

$$S_k(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \equiv \sum_{\mathbf{u}\in\mathbb{Z}^d} c^2_{g,\mathbf{n}_k}(\mathbf{u}) \left\{ c_X(\mathbf{u}; \boldsymbol{\theta}_1) - c_X(\mathbf{u}; \boldsymbol{\theta}_2) \right\}^2, \quad \forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta^2. \tag{19}$$

We remind the reader that the sums in (18) and (19) are *de facto* finite (although they might diverge to infinity with k), due to the definition of $c_{g,\mathbf{n}}(\mathbf{u})$. We observe that the above definition depends on both the sequence of grids, from $c_{g,\mathbf{n}_k}(\mathbf{u})$, and on the model family, from $c_X(\mathbf{u})$. In the rest of this paper we shall say that a sequence of grids leads to SCC, if the model family this applies to is obvious from the context. In addition we define the notion of Highly Significant Correlation Contribution (HSCC), which will allow us to establish a convergence rate.

Definition 2 (Highly Significant Correlation Contribution). *A sequence of observed grids* $(\mathcal{J}_{\mathbf{n}_k}, g_k)_{k\in\mathbb{N}}$ *leads to Highly Significant Correlation Contribution for the model family* $\{f(\,\cdot\,; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$

- *if it leads to Significant Correlation Contribution,*

- *if the covariance function is differentiable with respect to the parameter vector, and in particular, the quantity* $\min_{\mathbf{v}\in\mathbb{R}^p, \|\mathbf{v}\|=1} \sum_{\mathbf{u}\in\mathbb{Z}^d} c^2_{g,\mathbf{n}_k}(\mathbf{u}) \left(\sum_{j=1}^p v_j \left\{ \frac{\partial c_X}{\partial \boldsymbol{\theta}_j}(\mathbf{u}; \boldsymbol{\theta}) \right\} \right)^2$ *is asymptotically lower-bounded by a non-zero value, denoted* $S(\boldsymbol{\theta})$.

- *if the expected periodogram is twice differentiable with respect to the parameter vector, and such that its first and second derivatives are both upper-bounded in absolute value by a constant denoted* $M_{\partial\theta^2} > 0$.

Note that a necessary and more intuitive condition for the second item of the above definition is that for all $j = 1\ldots, d$, $\sum_{\mathbf{u}\in\mathbb{Z}^d} c^2_{g,\mathbf{n}_k}(\mathbf{u}) \frac{\partial c_X}{\partial \boldsymbol{\theta}_j}(\mathbf{u}; \boldsymbol{\theta})^2$ be lower-bounded by a positive value. Broadly speaking, the first part of (18) is required so that information grows fast enough. It can be compared to necessary conditions of decaying covariances in laws of large numbers, with the additional requirement of accounting for sampling when considering spatial data. Note that the first part of (18) is obviously satisfied if the sample covariance sequence is assumed square summable and the number of observations grows infinite.

The second part of (18) ensures that the expected periodograms for any two parameter vectors of the parameter set remain *asymptotically distant in terms of $\mathcal{L}_2$ norm.* In Lemma 11 in Section 5, we show how this transfers to the expectation of the likelihood function, ensuring that it attains its minimum at the true parameter vector uniquely. Then in Lemma 14 we show that the likelihood function converges uniformly in probability to its expectation over the parameter set, as long as the first part of (18) is satisfied. This all together will eventually lead to the consistency of our inference procedure, which is the result of Theorem 1. Hence the second part of (18) is required to ensure that the sampling allows to distinguish parameter vectors based on the expectation of our approximate likelihood function. To provide further understanding, we shall now consider some general cases and specific examples with respect to this definition.

4.3 General cases and example

Definition 1 extends the definition of SCC provided by Guillaumin et al. (2017) for time series in two ways. First, it provides a generalization for spatial data with the notable difference that spatial *sampling* is more complex than in time. Indeed, one needs to not only account for the frequency of the sampling but also for the direction. Secondly, even in dimension one, the version provided by Guillaumin et al. (2017) implies the version provided here, while the reverse is not always true—thus relaxing the assumptions required for consistency. In the second part of equation (18), we do not require to observe a specific finite set of lags that will allow identification of the parameters as by Guillaumin et al. (2017). We now provide more intuition about SCC through general cases and then provide two examples.

4.3.1 General cases

Under standard sampling conditions, SCC takes a simpler form, as we show through the two following lemmata.

Lemma 6 (SCC for full grids). *If we observe a sequence of full rectangular grids that grow unbounded in all directions (i.e., $n_i \to \infty$, $i = 1, \ldots, d$), SCC is then equivalent to the standard assumption that for any two distinct parameter vectors $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, the measure of the set $\{\boldsymbol{\omega} \in \mathcal{T}^d : f_{X,\boldsymbol{\delta}}(\boldsymbol{\omega}; \boldsymbol{\theta}_1) \neq f_{X,\boldsymbol{\delta}}(\boldsymbol{\omega}; \boldsymbol{\theta}_2)\}$ is positive.*

Proof. Please see the Supplementary Material. □

Importantly, we do not require that the growth happens with the same rate in all directions. We do require that grids grow unbounded in all directions to obtain this equivalence when we have no further knowledge on the functional form of the spectral densities. However, in many practical cases, such as that of an isometric exponential covariance function, our results still hold if the grid grows unbounded in one direction rather than all. Another important case for practical applications is that of a fixed shape of observations that grows unbounded, which is the subject of the following lemma.

Lemma 7 (Fixed shape of observations). *Consider a fixed shape defined by a function $\Xi : [0,1]^d \mapsto \{0, 1\}$, and let $g_{k,\mathbf{s}} = \Xi(\mathbf{s} \circ \mathbf{n}_k^{-1}), \forall \mathbf{s} \in \mathcal{J}_{\mathbf{n}_k}, \forall k \in I\!N$. If the grids grow unbounded in all directions, and if the interior of the support of Ξ is not empty, then SCC is again equivalent to the condition stated in Lemma 6 on the parametric family of spectral densities.*

Proof. Please see the Supplementary Material. □

In Section 6.2 we provide a simulation study for the particular case of a circular shape of observations. Finally, from a frequency-domain point of view, the second part of SCC can be understood according to the following lemma.

9

Lemma 8. *The second part of SCC is equivalent to*

$$S_k(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \int_{\mathcal{T}^d} \left| \int \mathcal{F}_{\mathbf{n}_k}(\boldsymbol{\omega}') \left\{ f_X(\boldsymbol{\omega}' - \boldsymbol{\omega}; \boldsymbol{\theta}_1) - f_X(\boldsymbol{\omega}' - \boldsymbol{\omega}; \boldsymbol{\theta}_2) \right\} d\boldsymbol{\omega}' \right|^2 d\boldsymbol{\omega} > 0.$$

Proof. This comes as a consequence of Lemma 3 and standard Fourier theory. □

Most importantly, note that in general SCC requires more than the basic requirement that for two distinct parameters, the expected periodograms for the sequence of grids should be non-equal, and this is to correctly account for missing data mechanisms and their impact on consistency. We will now study SCC for an isometric model family to provide some intuition and show the generality of this framework.

4.3.2 Example

We consider a separable exponential covariance function ($d = 2$ here) defined by

$$c_X(\mathbf{u}) = \sigma^2 \exp\left(-\rho_1^{-1}|u_1|\right) \exp\left(-\rho_2^{-1}|u_2|\right), \quad \mathbf{u} \in \mathbb{R}^2. \tag{20}$$

In the scenario where we sample along one axis only, it is clear that the second part of SCC fails as the range parameter along the other axis cannot be identified from the data. Interestingly, the second part of SCC will be satisfied for this particular model and for a full rectangular grid as long as $n_1 \geq 2$ and $n_2 \geq 2$. The first part of SCC is valid as long as the sample size grows to infinity, since the sample covariance function is square summable. Therefore for this model family, SCC is satisfied if and only if $n_1 \geq 2$ and $n_2 \geq 2$ and $n_1 n_2$ goes to infinity. It is also worth observing that under those conditions, the convergence rate of our estimator will be $\mathcal{O}\left((n_1 n_2)^{-1/2}\right)$ (see Theorem 2), irrespective of the ratio $\frac{n_1}{n_2}$, which in particular is allowed to converge to zero or infinity.

These two examples show the flexibility of SCC compared to standard assumptions. They show that the two parts of SCC are complimentary and help understand their role in establishing consistency. The second part is required to ensure identifiabilty of the parameter vector from the expected periodogram. The first part of SCC is required to ensure that the observed periodogram becomes uncorrelated at distinct frequencies.

4.3.3 Application to randomly missing data

Our extended definition of SCC can be applied to the scenario where data are missing at random, on the condition that the randomness scheme for the missing data is independent from that of the observed process. For such applications we shall say that a sequence of grids leads to SCC almost surely if (18) is satisfied almost surely under the probability that defines the missingness scheme. If a sequence of grids leads to SCC almost surely, it is easy to verify that all our consistency results derived in Section 5 still hold.

A simple application of these considerations is that where each point of a rectangular grid is observed or missed according to a Bernoulli random variable (with a positive probability of being observed), independently of other points of the grid, and independently of the observed process.

5 Theory

In this section, we provide the proof of our estimator's consistency, derive its rate of convergence and its asymptotic distribution. We assume the following set of assumptions holds in order to establish consistency.

Assumption 1 (Consistency assumptions).

1. *The parameter set Θ is compact.*

2. *The aliased spectral density $f_{X,\delta}(\boldsymbol{\omega};\boldsymbol{\theta}), \boldsymbol{\omega} \in \mathcal{T}^d, \boldsymbol{\theta} \in \Theta$ is bounded above by $f_{\delta,max} < \infty$ and below by $f_{\delta,min} > 0$. Additionally, $f_{X,\delta}(\boldsymbol{\omega};\boldsymbol{\theta})$ admits a derivative with respect to the parameter vector θ, which is upper-bounded by $M_{\partial\theta}$.*

3. *The sequence of observation grids leads to SCC for the considered model family.*

4. *The modulation $g_{\mathbf{s}}$, $\mathbf{s} \in \mathbb{Z}^d$, takes its values in the interval $[0,1]$.*

Two main asymptotic frameworks coexist in spatial data analysis, infill asymptotics and growing-domain asymptotics (Zhang and Zimmerman, 2005). We study our estimator within the latter framework, which we consider most plausible for finite-resolution remote-sensing observations, imposing that the sample size goes to infinity (through our SCC assumption) while having fixed $\boldsymbol{\delta}$. Our set of assumptions is standard, except for SCC, which generalizes the standard assumption of a fully-observed rectangular grid associated with the requirement that two distinct parameter vectors map to two spectral densities which are distinct on a Lebesgue set of non-zero measure.

Theorem 1 (Consistency). *Under the set of Assumptions 1, the sequence of estimates $\hat{\boldsymbol{\theta}}_k$ defined by (7) converges in probability to the true parameter vector $\boldsymbol{\theta}$ as the observational domain diverges.*

This result holds for a wide class of practical applications, as

- we do not require the rectangular grid to be fully observed. We allow for a wide class of observational domains, as long as SCC is satisfied.

- we do not require the grid to grow at the same rate along all dimensions. Classical frequency-domain results make use of the fact that the multilevel Block Toeplitz with Toeplitz Blocks covariance matrix has its eigenvalues distributed as the spectral density. However this result only holds under the assumption that the sampling grid grows at the same rate along all dimensions.

We shall prove Theorem 1 in a series of steps, but start by introducing some additional notation.

5.1 Additional notation

The vector of the values taken by the process on the rectangular grid $\mathcal{J}_\mathbf{n}$ is denoted $\mathbf{X} = [X_0, \ldots, X_{|\mathbf{n}|-1}]^T$, where points are ordered into a vector according to the colexicographical order. Therefore in dimension $d = 2$, $X_0, \ldots, X_{n_1-1}$ are values from the first row of $\mathcal{J}_\mathbf{n}$, $X_{n_1}, \ldots, X_{2n_1-1}$ are values from the second row, and so on. Similarly we denote $\mathbf{g}$ the vector of the values taken by the modulation function on $\mathcal{J}_\mathbf{n}$, with points ordered in the same way. We also denote by $\mathbf{s}_0, \ldots, \mathbf{s}_{|\mathbf{n}|-1}$ the locations of the grid ordered according to the same order, such that $X_0 = X(\mathbf{s}_0)$, $X_1 = X(\mathbf{s}_1)$, etc.

We also denote by G the diagonal matrix with elements taken from $\mathbf{g}$, such that the vector corresponding to the observed random field (rather than $\mathbf{X}$ which corresponds to the random field on the rectangular grid $\mathcal{J}_\mathbf{n}$) is given by the matrix product $G\mathbf{X}$.

Finally, for any vector $\mathbf{V}$ we shall denote by $\|\mathbf{V}\|_p$ its $\mathcal{L}_p$ norm (in particular $\|\cdot\|_2$ is the Euclidean norm), and for any $p \times p$ matrix A, $\|A\|$ shall denote the spectral norm, i.e., the $\mathcal{L}_2$-induced norm,

$$\|A\| = \max_{\mathbf{V} \in \mathrm{I\!R}^p, \mathbf{V} \neq \mathbf{0}} \frac{\|A\mathbf{V}\|_2}{\|\mathbf{V}\|_2}. \tag{21}$$

We remind the reader that if H is a Hermitian matrix, since $\|H\mathbf{V}\|_2^2 = \mathbf{V}^* H^* H \mathbf{V} = \mathbf{V}^* H^2 \mathbf{V}$, the spectral norm of H is its spectral radius, i.e.,

$$\|H\| = \rho(H) \equiv \max\{|\lambda| : \lambda \text{ eigenvalue of } H\}.$$

5.2 Distributional properties of the periodogram

It is well known that the bias of the periodogram as an estimator of the spectral density is asymptotically zero (see, e.g., Koopmans, 1995). However, in dimension $d \geq 2$, the decay of the bias of the periodogram is known to be the dominant factor in terms of mean-squared error (Dahlhaus and Künsch, 1987). By directly fitting the expectation of the periodogram, rather than the spectral density, we circumvent this major pitfall of the Whittle likelihood for random fields. Having removed the effect of bias, we are left with studying the correlation properties of the periodogram. We show that the variance of a bounded linear combination of the periodogram at Fourier frequencies goes to zero. This is the result of Proposition 1, which we use later in Lemma 14 to prove that our likelihood function converges uniformly in probability to its expectation.

Proposition 1 (Variance of linear functionals of the periodogram). *Let $a_k(\boldsymbol{\omega})$ be a family of functions with support $\mathcal{T}^d$, indexed by $k \in \mathrm{I\!N}$, and uniformly bounded in absolute value. Then,*

$$\mathrm{var}\left\{|\mathbf{n}_k|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} a_k(\boldsymbol{\omega}) I_{\mathbf{n}_k}(\boldsymbol{\omega})\right\} = \mathcal{O}\left\{\frac{\sum_{\mathbf{u} \in \mathbb{Z}^d} c_X(\mathbf{u})^2 c_{g,k}(\mathbf{u})}{\sum g_\mathbf{s}^2}\right\}. \tag{22}$$

Proof. Please see the Supplementary Material. $\square$

We note that in the simple case where we observe a full rectangular grid with no tapering we have $\sum g_{\mathbf{s}}^2 = |\mathbf{n}|$.

5.3   Lemmata required for Theorem 1

We provide all the proofs of this section in the Supplementary Material. To establish consistency we introduce some specific notation for the expectation of our quasi-likelihood,

$$\widetilde{\ell}_{\mathbf{n}}(\boldsymbol{\gamma}) = \mathrm{E}_{\boldsymbol{\theta}}\left\{\ell_{\mathbf{n}}(\boldsymbol{\gamma})\right\} = |\mathbf{n}|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}}} \left\{ \log \overline{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\gamma}) + \frac{\overline{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\theta})}{\overline{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\gamma})} \right\}, \quad \forall \mathbf{n} \in (\mathrm{I\!N}^+)^d \setminus \{\mathbf{0}\}, \forall \boldsymbol{\gamma} \in \Theta, \tag{23}$$

which we shall regard as a function of $\boldsymbol{\gamma}$. The following lemma relates the minimum of that function to the true parameter vector (with no uniqueness property as of now).

Lemma 9 (Minimum of the expected quasi-likelihood function). *The expected likelihood function attains its minimum at the true parameter value, i.e,*

$$\widetilde{\ell}_{\mathbf{n}}(\boldsymbol{\theta}) = \min_{\boldsymbol{\gamma} \in \Theta} \widetilde{\ell}_{\mathbf{n}}(\boldsymbol{\gamma}). \tag{24}$$

We shall also make repeated use of the following lemma.

Lemma 10 (Lower and upper bounds on the expected periodogram). *The expected periodogram satisfies, for all parameter vector $\boldsymbol{\gamma} \in \Theta$, and at all wave-numbers $\boldsymbol{\omega} \in \mathcal{T}^d$, for any $\mathbf{n} \in \mathrm{I\!N}^d$,*

$$f_{\delta,min} \leq \overline{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\gamma}) \leq f_{\delta,max}.$$

We now provide additional lemmata which are key to proving the consistency of our maximum quasi-likelihood estimator. Lemma 11 states that the expected likelihood value at a parameter vector distinct from the true parameter value is asymptotically bounded away from the expected likelihood at the true parameter value. This comes as a consequence of the second part of SCC and the upper-bound on the spectral densities of the model family.

Lemma 11 (Identifiability from the expected likelihood function). *Let $\boldsymbol{\gamma} \in \Theta$ distinct from $\boldsymbol{\theta}$. Then,*

$$\underline{\lim}_{k\to\infty} \left| \widetilde{\ell}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - \widetilde{\ell}_{\mathbf{n}_k}(\boldsymbol{\theta}) \right| > 0, \tag{25}$$

where $\underline{\lim}_{k\to\infty}$ denotes the limit inferior as k goes to infinity.

Lemma 12 now states a form of regularity of our expected likelihood functions.

Lemma 12. *Let $\boldsymbol{\gamma} \in \Theta$ and let $(\boldsymbol{\gamma}_k)_{k\in\mathrm{I\!N}}$ be a sequence of parameter vectors that converges to $\boldsymbol{\gamma}$. Then,*

$$\widetilde{\ell}_{\mathbf{n}_k}(\boldsymbol{\gamma}_k) - \widetilde{\ell}_{\mathbf{n}_k}(\boldsymbol{\gamma}) \longrightarrow 0, \quad (k \longrightarrow \infty). \tag{26}$$

Lemma 13. *Let $\boldsymbol{\gamma}_k \in \Theta^{\mathrm{I\!N}}$ be a sequence of parameter vectors such that $\widetilde{\ell}_{\mathbf{n}_k}(\boldsymbol{\gamma}_k) - \widetilde{\ell}_{\mathbf{n}_k}(\boldsymbol{\theta})$ converges to zero as k goes to infinity. Then $\boldsymbol{\gamma}_k$ converges to $\boldsymbol{\theta}$.*

And finally, the following lemma helps us understand how the likelihood function, as a random element, behaves with regard to the expected likelihood function.

Lemma 14 (Uniform convergence in probability of the likelihood function)**.** *The likelihood function $\ell_{\mathbf{n}_k}(\cdot)$ converges uniformly in probability to $\widetilde{\ell}_{\mathbf{n}_k}(\cdot)$ over the parameter set Θ as k goes to infinity.*

With these lemmata we have all the necessary results to establish Theorem 1. This theorem is important as it establishes the consistency of our estimator under a very wide range of sampling schemes and model familites. We contrast our results with those of Dahlhaus and Künsch (1987), Guyon (1982), as well as Fuentes (2007). The insight from Theorem 1, as compared to the insight of the need for tapering provided by Dahlhaus and Künsch (1987) is clear. The aim of this paper is to balance computational tractability with estimation performance. Very standard assumptions allow us to still derive the results required for estimation.

5.4   Convergence rate and asymptotic normality

We now study the convergence rate and asymptotic distribution of our estimates within the increasing-domain asymptotics framework. We establish a convergence rate in the general framework of SCC, and additionally establish asymptotic normality in the scenario of a full grid. We first need to understand better the behaviour of quantities of the form $|\mathbf{n}|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} w_k(\boldsymbol{\omega}) I_{\mathbf{n}}(\boldsymbol{\omega})$, for some weights w_k. In Proposition 1, we showed that under mild conditions, their variance was vanishing, with a rate driven by the number of observed points. In Proposition 2, and under the assumption of a full grid, by writing this quantity as a quadratic form in the random vector $\mathbf{X}$, and by extending a result by Grenander and Szegö (1958), we show that this quantity is asymptotically normally distributed, under mild conditions on the family of functions $w_k(\cdot)$. Before getting there, we need the following intermediary result, which extends a standard result for Toeplitz matrices to their multidimensional counterpart, Block Toeplitz with Toeplitz Block matrices.

Lemma 15 (Upper bound on the spectral norm of the covariance matrix)**.** *In the case of a full grid, the spectral norm of $C_{\mathbf{X}}$ and that of its inverse are upper-bounded according to*

$$\|C_{\mathbf{X}}\| \leq f_{\delta,max}, \quad \|C_{\mathbf{X}}^{-1}\| \leq f_{\delta,min}^{-1}.$$

Proof. Please see the Supplementary Material. □

Proposition 2 (Asymptotic normality of bounded linear combinations of the periodogram)**.** *Let $X(\mathbf{s})$, $\mathbf{s} \in I\!R^d$ be a homogeneous Gaussian process observed on the lattice $\mathcal{J}_{\mathbf{n}}$. Let $w_k(\cdot), k \in I\!N$ be a family of real-valued functions defined on $\mathcal{T}^d$ bounded above and below by two constants, denoted $M_W, m_W > 0$ respectively. Additionally, assume $g_s = 1, \forall \mathbf{s} \in \mathcal{J}_{n_k}$. Then $|\mathbf{n}|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} w_k(\boldsymbol{\omega}) I_{\mathbf{n}}(\boldsymbol{\omega})$ is asymptotically normally distributed.*

Proof. Please see the Supplementary Material. □

Before finally establishing our convergence rates, as well as the asymptotic normality in the case of full grids, we require one additional set of assumptions.

Assumption 2 (Assumptions for convergence rate and asymptotic normality)**.**

1. *The interior of Θ is non-null and the true length-p parameter vector $\boldsymbol{\theta}$ lies in the interior of Θ.*

2. *The sequence of grids leads to HSCC.*

The following lemma relates HSCC to the minimum eigen value of the expectation of the Hessian matrix of $l(\boldsymbol{\theta})$ at the true parameter vector.

Lemma 16. *Under HSCC, the minimum eigenvalue of the expectation of the Hessian matrix (with respect to the parameter vector) at the true parameter, given by*

$$\left(|\mathbf{n}_k|^{-1} \sum_{\boldsymbol{\omega} \in \Omega} \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})^{-2} \nabla_\theta \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta}) \nabla_\theta \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})^T \right), \tag{27}$$

is lower-bounded by $S(\boldsymbol{\theta})$, which was defined in Definition 2.

Proof. This can be established by a direct adaptation of Lemma 7 of Guillaumin et al. (2017). □

Theorem 2 (Convergence rate and asymptotic normality of estimates)**.** *Under the sets of Assumptions 1 and 2, our estimate converges with rate*

$$r_k = \left(\sum g_{\mathbf{s}}^2 \right)^{-1/2} \left(\sum_{u \in \mathbb{Z}} c_{g, \mathbf{n}_k}(\mathbf{u}) c_X(\mathbf{u})^2 \right)^{1/2}. \tag{28}$$

In the case of a full grid, i.e. $g_{\mathbf{s}} = 1, \forall \mathbf{s} \in \mathcal{J}_{\mathbf{n}}$, the estimate $\hat{\boldsymbol{\theta}}$ is asymptotically normally distributed and converges with rate $|\mathbf{n}_k|^{-1/2}$.

Proof. Please see the Supplementary Material. □

In the case of a full grid, having established asymptotic normality allows to derive confidence intervals based on the estimation of standard errors, which is the topic of the following section.

5.5 Estimating standard errors

We now seek to derive how to estimate the standard error of $\widehat{\boldsymbol{\theta}}$. Using equations (41) and (42) from the Supplementary Material, we obtain an approximation for the variance of $\widehat{\boldsymbol{\theta}}$ in the following proposition.

Proposition 3 (Exact form of the variance)*. The covariance matrix of the quasi-likelihood estimator takes the form of*

$$\text{var}\left\{\widehat{\boldsymbol{\theta}}\right\} \approx \boldsymbol{\mathcal{H}}^{-1}(\boldsymbol{\theta})\text{var}\left\{\nabla\ell_M(\boldsymbol{\theta})\right\}\boldsymbol{\mathcal{H}}^{-1}(\boldsymbol{\theta}), \tag{29}$$

with the covariance matrix of the score taking the form of

$$\text{cov}\left\{\frac{\partial\ell_M(\boldsymbol{\theta})}{\partial\theta_p}, \frac{\partial\ell_M(\boldsymbol{\theta})}{\partial\theta_q}\right\} = |\mathbf{n}|^{-2}\sum_{\boldsymbol{\omega}_1\in\Omega_\mathbf{n}}\sum_{\boldsymbol{\omega}_2\in\Omega_\mathbf{n}}\frac{\partial\overline{I}_\mathbf{n}(\boldsymbol{\omega}_1;\boldsymbol{\theta})}{\partial\theta_p}\frac{\partial\overline{I}_\mathbf{n}(\boldsymbol{\omega}_2;\boldsymbol{\theta})}{\partial\theta_q}\frac{\text{cov}\left\{I_\mathbf{n}(\boldsymbol{\omega}_1), I_\mathbf{n}(\boldsymbol{\omega}_2)\right\}}{\overline{I}_\mathbf{n}^2(\boldsymbol{\omega}_1;\boldsymbol{\theta})\overline{I}_\mathbf{n}^2(\boldsymbol{\omega}_2;\boldsymbol{\theta})}. \tag{30}$$

The computation that appears in (30) does not scale well for large grid sizes, as it scales like $|\mathbf{n}|^2$. We instead propose a Monte Carlo implementation to speed this up. The dominant terms in (30) correspond to $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_2$. We approximate the sum over the rest of the terms, i.e., our approximation takes the form:

$$\text{cov}\left\{\frac{\partial\ell_M(\boldsymbol{\theta})}{\partial\theta_p}, \frac{\partial\ell_M(\boldsymbol{\theta})}{\partial\theta_q}\right\} = \frac{1}{|\mathbf{n}|^2}\sum_{\boldsymbol{\omega}_1\in\Omega_\mathbf{n}}\left\{\frac{\partial\overline{I}_\mathbf{n}(\boldsymbol{\omega}_1;\boldsymbol{\theta})}{\partial\theta_p}\frac{\partial\overline{I}_\mathbf{n}(\boldsymbol{\omega}_1;\boldsymbol{\theta})}{\partial\theta_q}\frac{\text{var}\left\{I_\mathbf{n}(\boldsymbol{\omega}_1)\right\}}{\overline{I}_\mathbf{n}^4(\boldsymbol{\omega}_1;\boldsymbol{\theta})}\right\}$$
$$+ \frac{n^2 - 2n + 1}{M|\mathbf{n}|^2}\sum_{i=1\ldots M}\frac{\partial\overline{I}_\mathbf{n}(\boldsymbol{\omega}_{1,i};\boldsymbol{\theta})}{\partial\theta_p}\frac{\partial\overline{I}_\mathbf{n}(\boldsymbol{\omega}_{2,i};\boldsymbol{\theta})}{\partial\theta_q}\frac{\text{cov}\left\{I_\mathbf{n}(\boldsymbol{\omega}_{1,i}), I_\mathbf{n}(\boldsymbol{\omega}_{2,i})\right\}}{\overline{I}_\mathbf{n}^2(\boldsymbol{\omega}_{1,i};\boldsymbol{\theta})\overline{I}_\mathbf{n}^2(\boldsymbol{\omega}_{2,i};\boldsymbol{\theta})},$$

where the $\boldsymbol{\omega}_{1,i}, \boldsymbol{\omega}_{2,i}, i = 1\ldots M$ are uniformly and independently sampled from the set of Fourier frequencies $\Omega_\mathbf{n}$ under the requirement $\boldsymbol{\omega}_{1,i} \neq \boldsymbol{\omega}_{2,i}$. Note that if tapering is used, one should consider a few coefficients near the main diagonal in the above approximation, as tapering generates strong short-range correlation in the frequency domain.

The covariances of the periodogram at two distinct Fourier frequencies can be approximated by Riemann approximation of the two integrals that appear in the expression below, before taking squared absolute values and summing,

$$\text{cov}\left\{I_\mathbf{n}(\boldsymbol{\omega}_{1,i}), I_\mathbf{n}(\boldsymbol{\omega}_{2,i})\right\} = |\mathbf{n}|^{-1}\left(\left|\int_{-\pi}^{\pi}\widetilde{f}(\lambda)\mathcal{D}_\mathbf{n}(\lambda - \boldsymbol{\omega}_{1,i})\mathcal{D}_\mathbf{n}^*(\lambda - \boldsymbol{\omega}_{2,i})\,d\lambda\right|^2\right.$$
$$\left.+ \left|\int_{-\pi}^{\pi}\widetilde{f}(\lambda)\mathcal{D}_\mathbf{n}(\lambda - \boldsymbol{\omega}_{1,i})\mathcal{D}_\mathbf{n}^*(\lambda + \boldsymbol{\omega}_{2,i})d\lambda\right|^2\right), \quad i = 1,\ldots, M.$$

In the above, $\widetilde{f}$ is the following approximation to the spectral density, which can be computed by a DFT,

$$\widetilde{f}(\lambda) = \sum_{\mathbf{u}\in\prod_{i=1}^d[-(n_i-1)\ldots(n_i-1)]}c_X(\mathbf{u};\boldsymbol{\theta})\exp(-i\lambda\cdot\mathbf{u}),$$

and $\mathcal{D}_\mathbf{n}(\lambda)$ is the non-centred *modified* (due to the modulation $g_\mathbf{s}$) Dirichlet kernel of order $\mathbf{n}$ given by

$$\mathcal{D}_\mathbf{n}(\lambda) = \sum_{\mathbf{s}\in\mathcal{J}_\mathbf{n}}g_\mathbf{s}\exp(i\lambda\cdot\mathbf{s}),$$

16

where for clarity we omit the dependence on the modulation $g_{\mathbf{s}}$ in the notation. Finally we compute the derivatives of $\bar{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\theta})$ as follows,

$$\nabla_{\boldsymbol{\theta}} \bar{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\theta}) = \sum_{u \in \mathbb{Z}^d} \nabla_{\boldsymbol{\theta}} \bar{c}_X(\mathbf{u}; \boldsymbol{\theta}) \exp(-i\boldsymbol{\omega} \cdot \mathbf{u}). \tag{31}$$

6 Simulation studies and application to the study of planetary topography

In this section we present simulation studies and an application to the study of Venus' topography that demonstrate the performance of the debiased spatial Whittle estimator. The simulations presented in Section 6.1 address the estimation of the range parameter of a Matérn process, whose slope parameter is known, observed over a full rectangular grid. These simulations corroborate our theoretical results on the optimal convergence rate of our estimator despite edge effects, in contrast to the standard Whittle method. Our second simulation study in Section 6.2 shows how our estimation procedure extends the computational benefits of frequency-domain methods to non-rectangular shapes of data, where we compare parameter estimates with those of Guinness and Fuentes (2017) in the scenario of a circular shape of observations. In Section 6.3 we estimate the parameters of a simulated Matern process sampled according to a real-world sampling scheme of terrestrial ocean-floor topography (GEBCO Bathymetric Compilation Group, 2019) with approximately 72% missing data. Finally, in Section 6.4 we demonstrate the performance of the debiased spatial Whittle estimator when applied to topographical datasets obtained from Venus (Rappaport et al., 1999).

6.1 Estimation from a fully-observed rectangular grid of data

We simulate from the isotropic Matérn model family, which corresponds to the following covariance function,

$$c_X(\mathbf{u}) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|\mathbf{u}\|}{\rho} \right)^{\nu} \cdot K_{\nu} \left(\sqrt{2\nu} \frac{\|\mathbf{u}\|}{\rho} \right), \tag{32}$$

where $K_{\nu}(x)$ is a Bessel function of the second kind. We consider the problem of estimating the range parameter ρ, which is fixed to 10 units, while the amplitude $\sigma^2 = 1$ and the slope parameter $\nu \in \{\frac{1}{2}, \frac{3}{2}\}$ are fixed and known. Inference is achieved from simulated data on two-dimensional rectangular grids of increasing sizes, specifically $\{2^s : s = 4, \cdots, 8\}$ in each dimension.

We implement four inference methods:

(1) The debiased Whittle method, i.e., the estimate derived from (7);

(2) The debiased Whittle method combined with a taper, specifically the estimate derived from (7) with $g_{\mathbf{s}}$ proportional to a Hanning taper;

(3) The standard Whittle likelihood, i.e., estimators obtained by replacing $\overline{I}_{\mathbf{n}}(\boldsymbol{\omega};\boldsymbol{\theta})$ with $f_X(\boldsymbol{\omega})$ in (5) and then minimizing (7);

(4) The standard Whittle likelihood combined with tapering using a Hanning taper, again derived from (7) fitting to $f_X(\boldsymbol{\omega})$.

For each configuration of the slope parameter and grid size, we report summary statistics corresponding to 1,000 independently realized random fields. We report bias, standard deviation and root mean-squared error for $\nu = 0.5$ and $\nu = 1.5$ in Figures 1 and 2, respectively.

We first observe that the rate of the Whittle likelihood (3) is very poor, due to its large bias. It appears that tapering (4) leads to improved convergence rates when $\nu = 1.5$, although bias remains. On the contrary, the rates of our proposed method (1) and its tapered version (2) seem to not curb down even with larger grid sizes. This concurs with the theoretical results on the rate of convergence provided in Section 5. This example demonstrates that our debiased Whittle method balances the need for computational and statistical efficiency with large data sets.

In Figure 3 we report the empirical distribution of each estimator obtained from the 1,000 independent inference procedures for $\nu = \frac{1}{2}$. The four panels (a), (b), (c) and (d) show the distribution of estimates from the four methods stated on the previous page. The first two panels, (a) and (b), are broadly unbiased with estimates centred at $\rho = 10$ that converge quickly. The standard Whittle method (c) has issues with underestimation, tending towards $\rho = 5$. This asymptotic bias is in large part due to aliasing not being accounted for combined with the relatively small value of $\nu = 0.5$; these effects are still present in the tapered estimates (d). As would be expected, in all four subplots the variance is decreasing with increasing sample size, at similar rates.

6.2 Estimation from a circular set of observations

In this section, we show how our debiased Whittle method extends to non-rectangular data. More specifically, we assume we only observe data within a circle with diameter 97 units. We consider the exponential covariance kernel given by

$$c_X(\mathbf{u}) = \sigma^2 \exp\left(-\frac{\|\mathbf{u}\|}{\rho}\right), \qquad \mathbf{u} \in \mathbb{R}^2, \tag{33}$$

where $\sigma^2 = 1$ is fixed and known and we estimate the range parameter ρ whose true value is set to 5 units. We note that the case of a growing circle satisfies SCC, according to Lemma 7, and hence leads to consistency and asymptotic normality of our estimator.

A total number of 1,200 independent simulations are performed. As a state-of-the-art baseline, we compare to a recent method proposed by Guinness and Fuentes (2017), which is an approximation of the circulant embedding method developed by Stroud et al. (2017). Stroud et al. (2017) propose an Expectation Maximization iterative procedure, where the observed sample is embedded onto a larger grid that makes the covariance matrix *Block*
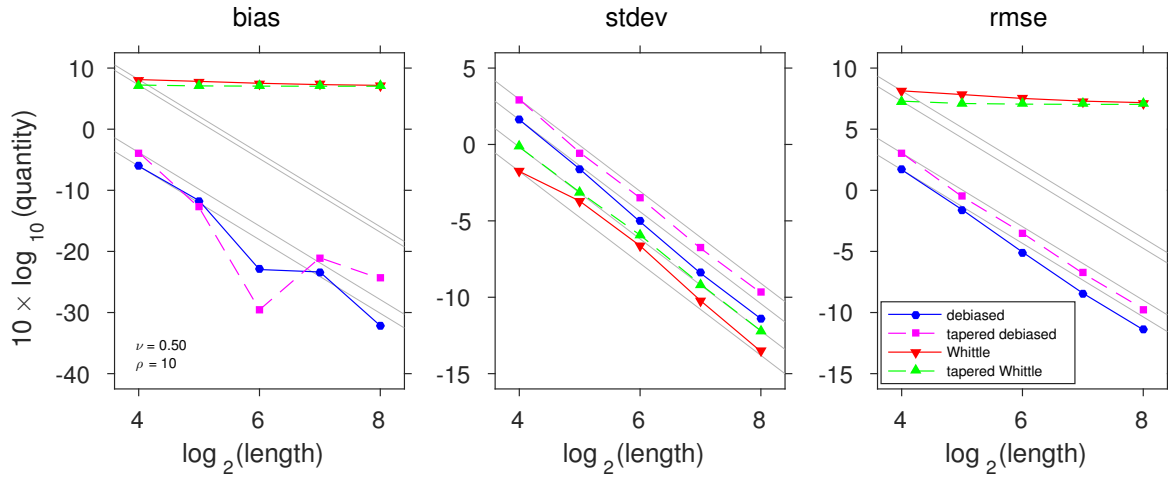
Figure 1: Bias, standard deviation, and root mean-squared error of estimates of the range parameter $\rho = 10$ of a Matérn process (32) with $\nu = \frac{1}{2}, \sigma^2 = 1$. The estimation method is identified by the line style, and gray lines functionally express the theoretical dependence on the square root of the sample size. The side length of the two-dimensional square grid is indicated by the horizontal axis, leading to a sample size of the length squared.
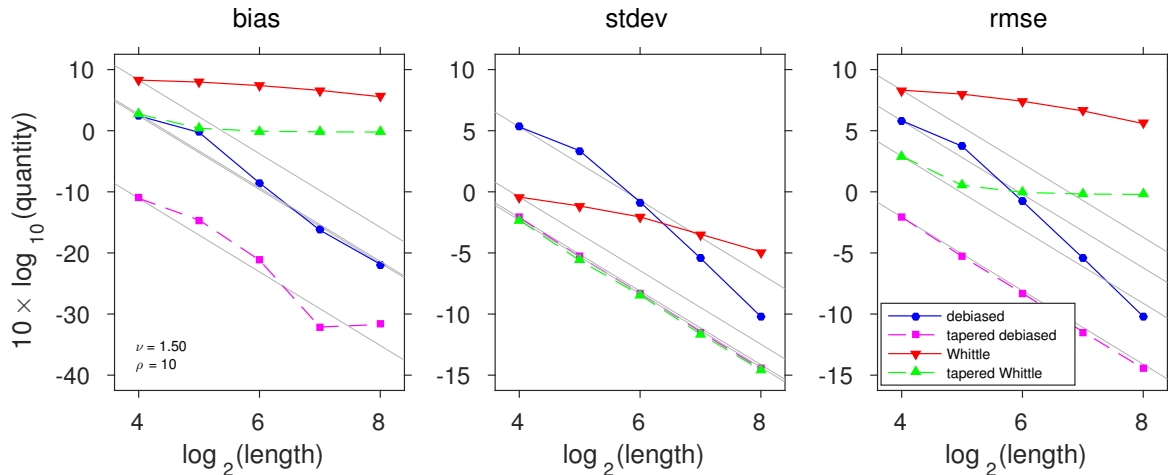

Figure 2: The same simulation setup as in Figure 1, but with $\nu = \frac{3}{2}$. This higher slope parameter is associated with smoother realizations, resulting in worsened edge effects. This illustrate how our method effectively addresses the edge effect issue even in that setting.
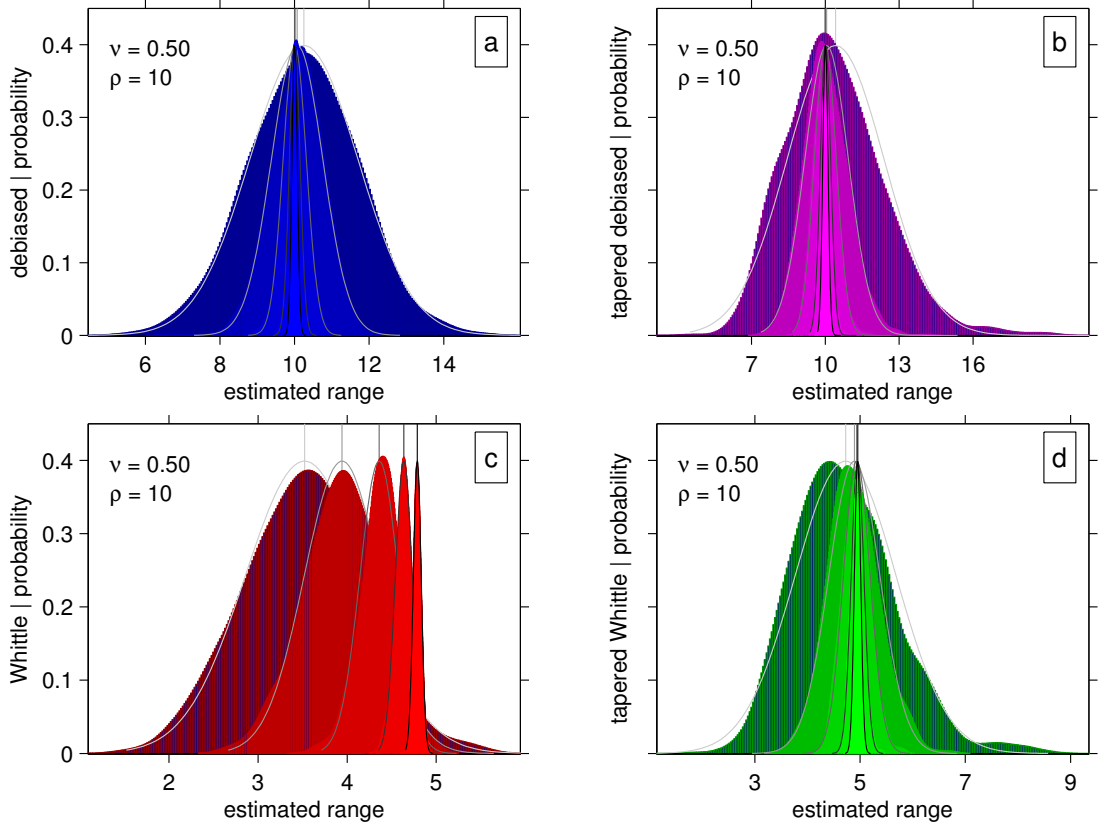
19

Figure 3: Nonparametric density estimates $\hat{\rho}$ of the estimated range parameter $\hat{\rho}$ ($\rho = 10$) for a Matérn random field (32), with $\sigma^2 = 1$ and $\nu = \frac{1}{2}$. The four subplots show different estimation methods of (a) debiased Whittle, (b) debiased Whittle with tapering, (c) standard Whittle, and (d) standard Whittle with tapering. The density estimate is shaded to reflect the size of the random field, with the darkest corresponding to total observations $|\mathbf{n}| = (2^4)^2$, and the shading incrementally taking a lighter colour for $|\mathbf{n}| = (2^5)^2, (2^6)^2, (2^7)^2, (2^8)^2$. Each density estimate is complemented by the best fitting Gaussian approximation as a solid black or fading gray line (black corresponds to $|\mathbf{n}| = (2^8)^2$ and the lightest gray to $|\mathbf{n}| = (2^4)^2$.)

Circulant with Circulant Blocks (BCCB), from where one can leverage the fast diagonalization procedure available for this class of matrices through the FFT algorithm. Guinness and Fuentes (2017) discuss that the size of the embedding grid is very large, making the imputations costly as well as the convergence of the iterative procedure slow. To address this they propose using a periodic approximation of the covariance function on an embedding grid which is much smaller than that required for the exact procedure. They show via simulations that using an embedding grid ratio of 1.25 along each axis leads to good approximations of the covariance function on the observed grid.

To implement the method developed by Guinness and Fuentes (2017), we use the code provided by the authors. We set a grid ratio of 1.25 to limit the computational cost, and implement the method with two choices of the number of imputations per iteration, $M = 1$ and $M = 20$. Each implementation is run for a number of 30 iterations for all samples.

Both our estimation method and that of Guinness and Fuentes (2017) are initialized with the estimates provided by the method proposed by Fuentes (2007). We show in Figure 4 (left panel) how debiased Whittle can achieve both computational and statistical efficiency. The 95 per cent confidence interval achieved by our estimate is similar to that obtained via the method of Guinness and Fuentes (2017) (M=1), however our method, despite also using an iterative maximization procedure, is significantly faster. As shown in Figure 4 (right panel), Guinness and Fuentes (2017) (M=20) leads to lower root mean squared error but requires higher computational time.
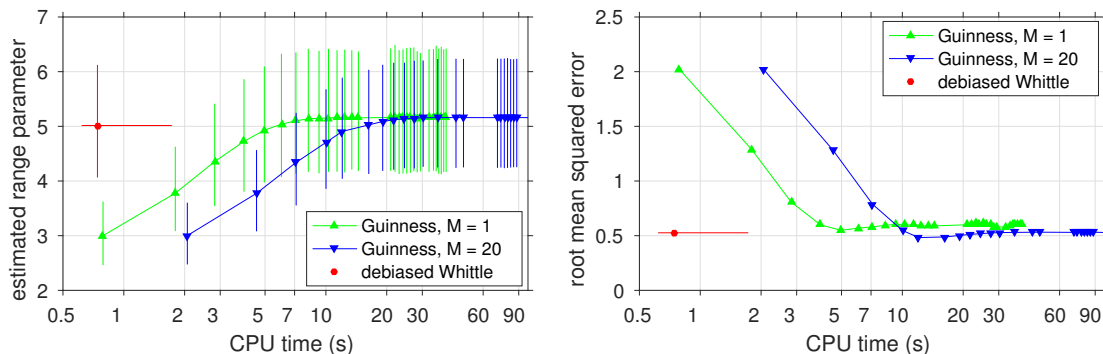

Figure 4: Mean and 95 per cent confidence intervals (left) and root mean-squared error (right) of estimates of the range parameter $\rho = 5$ of an exponential covariance model (33). Estimation is performed on a circular set of data with diameter 97 units. The converged estimates of the debiased Whittle method are compared to the iterated estimates of two implementations of Guinness and Fuentes (2017). The horizontal axis in both panels corresponds to the average computational time, as performed on an Intel(R) Core(TM) i7-7500U CPU 2.7–2.9 GHz processor.
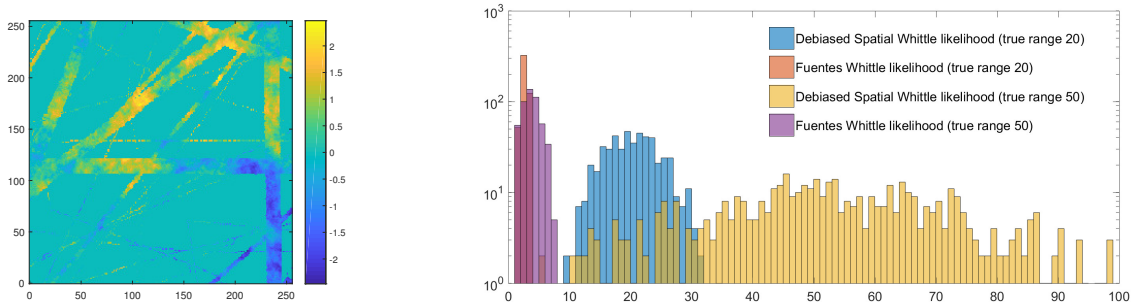
21

Figure 5: (Left) Simulated Matérn process with slope parameter 0.5 and range parameter 50 units, on a real-world sampling grid, with missing observations replaced by zeros. (Right) Histogram of estimates of the range parameter of a simulated Matérn process observed on the real-world grid shown in the left panel. We compare our proposed estimation method, the Debiased Spatial Whittle likelihood, to the method proposed by Fuentes (2007). The true value of the range is fixed to 20 or 50. Despite an increased variance due to the complex missing data patterns, our method is still able to produce a useful estimate of the range parameter, in comparison to the estimates produced by the method proposed by Fuentes, which was not built to address such large and complex patterns of missing data.

6.3 Application to a realistic sampling scheme of ocean-floor topography

In this simulation study we show that our estimator can address complex lower-dimensional sampling substructure. We apply it to the estimation of a Matrn process sampled on a real-world observation grid, characterized by a very large amount of missing data ($\approx 72\%$). We simulate two Matern processes, each with slope parameter 0.5 and with range 20 and 50 units respectively. The initial grid has size 1081×1081. For our study, we select a subgrid of size 256×256 with similar missingness properties to those of the whole grid. In Figure 5 we plot (left) a simulated Matérn process on that grid where missing observations have been replaced with zeros. We note the large amount of missing observations within the bounding rectangular grid, as well as its complex patterns (i.e. rather than a uniform missingness scheme). For this reason, the method proposed by Fuentes (2007) fails, while our method is still able to produce useful estimates, as shown in the right panel of Figure 5.

6.4 Application to the study of Venus' topography

In this section we apply our debiased spatial Whittle method to the study of Venus' topography. The motivation for modelling a planet's topography using a parametric covariance model such as the Matérn process is multifaceted. For instance, we may expect that the combination of the slope and range parameters will carry important information about the geomorphological process or age of formation of the observed topography, i.e. it is expected

Table 1: Estimates of the three parameters of a Matérn process (32)

	Patch 1			Patch 2			Patch 3			Patch 4		
Parameter:	σ	ν	ρ	σ	ν	ρ	σ	ν	ρ	σ	ν	ρ
Debiased Whittle	1.2	0.5	17.7	1.2	0.7	6.8	2.1	0.5	36.5	1.5	0.6	15.0
Standard Whittle	1.6	0.3	62.7	1.8	0.3	73.9	1.5	0.2	77.3	1.7	0.3	87.3
Tapered Whittle	2.0	0.4	52.0	1.7	0.2	80.6	1.2	0.2	88.1	1.9	0.4	83.7

Table 2: Percentage of increase in the exact likelihood value at the estimated parameter values from Table 1 in comparison to the minimal value obtained among the three methods.

	Patch 1	Patch 2	Patch 3	Patch 4
Debiased Whittle	60.60	104.80	91.60	48.40
Standard Whittle	0	16.10	0	0
Tapered Whittle	23.20	0	53.90	25.20

that those parameters will have an interpretable physical meaning. The slope parameter can be related to the smoothness of the topography, and the range parameter tells about the typical distance over which two observed portions are uncorrelated.

Building on the work of Eggers (2013), we have selected four patches of data (including that shown in Figure 6 which corresponds to Patch 3), each sampled regularly on a complete rectangular grid. We compare three estimation procedures: the debiased Whittle method, the standard Whittle method, and the standard Whittle method with tapering (again using a Hanning taper). Parameter estimates are reported in Table 1. We also compare the value of the exact likelihood function taken at the estimated parameters for each estimation method in Table 2. Specifically, if $\hat{\boldsymbol{\theta}}_M$ and $\hat{\boldsymbol{\theta}}_W$ respectively denote the estimates obtained via the debiased Whittle and standard Whittle procedure, we compare $l_E(\hat{\boldsymbol{\theta}}_M)$ and $l_E(\hat{\boldsymbol{\theta}}_W)$, with $l_E(\cdot)$ denoting the exact likelihood function (which is expensive to evaluate but only needs to be done once for each analyzed method). The results in Table 2 show a much better fit of the model corresponding to the parameters estimated via the debiased Whittle method, in comparison to the parameters estimated via either standard Whittle or tapered Whittle. The parameter estimates in Table 1 should be interpreted with care due to the challenges inherent in joint estimation of all three parameters of a Matérn covariance function (see, e.g., Zhang, 2004). However in all four patches we observe that the standard and tapered Whittle likelihood appear to overestimate the range while underestimating the smoothness, consistent with results found by Sykulski et al. (2019) for oceanographic time series.

To conclude our real-data analysis, we presented in Figure 6 a comparison of Patch 3 with three simulated samples, obtained using the Matérn model estimated using the debiased, standard and tapered Whittle methods respectively. This analysis supports the conclusion that the debiased Whittle method has found more appropriate parameter values for the model fit.
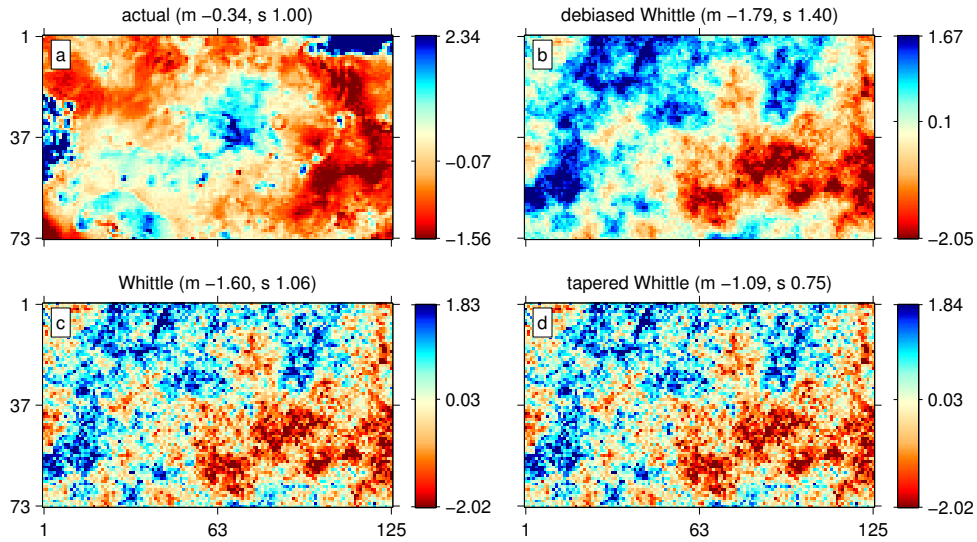
Figure 6:   (a) A realized random field from the topography of Venus; and simulated random fields from a Matérn model with parameters estimated using (b) debiased Whittle estimation, (c) standard Whittle estimation, and (d) standard Whittle estimation using a Hanning taper. Simulated random fields were obtained using the same random seed to facilitate comparison. Parameter values for each method are given in Table 1 (Patch 3) in Section 6.4. Sample means (m) and standard deviations (s) are in the titles. Color bars are marked at the 2.5th, 50th and 97.5th quantiles. Axis labels are in pixels.

7   Discussion

This paper has addressed the approximation of the likelihood of a Gaussian random field using the discrete Fourier transform. Key to understanding a random field is its spatial sampling; this can range from a spatial point process, to regular sampling with an irregular boundary, to observations missing at random to a fully sampled square regular grid. To maintain computational feasibility, this paper addresses the analysis of a regularly sampled random field, with potentially missing observations and an irregular boundary.

The Whittle likelihood uses the fast Fourier transform to achieve computational efficiency. The approximation is based on results for block-Toeplitz matrices (Tyrtyshnikov and Zamarashkin, 1998; Kazeev et al., 2013). This approximation is based on (growing-domain) asymptotics and arguments that equate the Gaussian non-diagonal quadratic form with another Gaussian, nearly diagonal, form. For time series this argument is relatively straightforward; but is somewhat more complex for spatial data in higher dimensions, where the bias becomes the dominant term, and the geometry of the sampling process leaves its imprint.

The bias of the periodogram, as an estimator of the spectral density (which drives subsequent bias), decreases with rate $\mathcal{O}\left(|\mathbf{n}|^{-1/d}\right)$ (Guyon, 1982; Dahlhaus and Künsch, 1987) in the case of a fully-observed rectangular lattice in d dimensions that grows at the same rate along all directions. A more precise result by Kent and Mardia (1996) shows that the approximation resulting from replacing the exact likelihood with the Whittle likelihood is driven by the size of the smallest side of the rectangular lattice. Dahlhaus (1983) proposed the use of tapering to remedy this issue.

Tapering alone can mitigate but not fully annihilate bias. We propose to replace the spectral density by the true expectation of the periodogram and thus perfectly remove bias. By introducing the notion of significant correlation contribution, we can also understand the mechanics of this process and draw a general framework of sampling schemes and model families for which our estimator is efficient.

For random fields with missing observations, Fuentes (2007) suggested to replace the missing points of a rectangular lattice with zeros as in (4) and correcting uniformly across frequencies for the amplitude of the periodogram, based on the ratio of the number of observed points to the total number of points in the grid. This only partly corrects for the bias of the periodogram that results from any non-trivial shape of the data, as frequencies are likely to not be affected uniformly by the sampling scheme.

Whittle estimation is often used when observations are non-Gaussian. This relaxation is less common for spatial observations. When studying non-Gaussian observations one can take two approaches; either limiting the effects of the non-Gaussianity on the variance of the estimator (Giraitis and Taqqu, 1999; Sykulski et al., 2019), or even permitting Whittle-type estimation based on higher order spectral moments, see e.g. Anh et al. (2007). It is also known that if infill asymptotics are considered (Bandyopadhyay and Lahiri, 2009), then the limiting distribution of the Fourier transform need not be Gaussian. Note that the aforementioned authors assumed completely random sampling of the field, which we do not assume, and such sampling leads to a "nugget-like effect" at frequency zero and

beyond.

We have assumed that the original observations are Gaussian, and whilst relaxing the growth of observations to be not necessarily matched in spatial directions, this leads to Gaussian estimators. We have carved out a large domain of spatial processes with non-regular sampling, and treating these effects fully with non-Gaussianity is too complex an issue to address in one instance.

Recently, Stroud et al. (2017) proposed a novel approach. Instead of approximating the multi-level Toeplitz covariance matrix of the rectangular lattice sample by a multi-level circulant matrix, one finds a larger lattice, termed an embedding, such that there exists a Block Circulant with Circulant Blocks (BCCB) matrix that is the covariance matrix of a Gaussian process on this extended lattice, and such that the covariance matrix of the real process is a submatrix of this extended matrix. One can then simulate efficiently the missing data on the extended lattice, and estimate the parameters of the models. This process can be iterated until a convergence criterion is met. This elegant method still suffers from computational issues, as the size of the embedding might be quite large. A solution suggested by Guinness and Fuentes (2017) is to use a circulant approximation of the covariance on a smaller rectangular lattice. The method is no longer exact, but Guinness and Fuentes (2017) showed via simulations that using small embeddings can in some cases provide a good compromise between statistical and computational efficiency.

In contrast, in this paper we have revisited the root cause of why the approximation of the likelihood may deteriorate, while still requiring any proposed bias elimination to result in a computationally competitive method. This bias elimination is "built in" by fitting the periodogram to $\bar{I}_{\mathbf{n}}(\boldsymbol{\omega}; \boldsymbol{\theta})$, which is the expected periodogram. This is in contrast to estimating the bias and removing it, as this procedure would typically increase variance, and might lead to empirical spectral density estimates that are negative.

We have thus proposed a bias elimination method that is data-driven, automated, and computationally practical for a number of realistic spatial sampling methods in any dimension. Our methods are robust to huge volumes of missing data, as evidenced by our theoretical investigations, as well as our practical simulation examples. As a result, our methodology is not only of great benefit for improved parameter estimation directly, but also has knock-on benefits in, for example, the problem of prediction. Here a huge number of methods exist and there is some debate as to which are most practically useful (Heaton et al., 2019), however the broader point is that many of these methods are based on Matérn covariance kernels—meaning our methods, which we have shown to greatly improve Matérn parameter estimation, can be naturally incorporated to improve the performance of spatial methods for prediction. Quantifying this benefit in a range of settings is a natural line of further investigation.

Within parameter estimation, there are a number of large outstanding challenges which are nontrivial extensions and merit further investigation as standalone pieces of work: 1) Extensions to irregularly sampled process on non-uniform grids; 2) Extensions to non-Gaussian random fields; and 3) Extensions to multivariate processes. In each case the impact on the Fourier Transform and the expected periodogram needs to be carefully han-

dled to properly account for the bias of naively using basic Whittle-type approximations—however we expect large improvements are possible both in terms of bias reduction (vs standard Whittle methods where edge effect contamination will increase) and in terms of computational speed (vs exact likelihood and other pseudo-likelihoods which will become increasingly intractable as assumptions are relaxed).

References

Anh, V. V., Leonenko, N. N., and Sakhno, L. M. (2007), "Minimum contrast estimation of random processes based on information of second and third orders," *J. Stat. Plan. Inf.*, 137(4), 1302–1331.

Anitescu, M., Chen, J., and Stein, M. L. (2017), "An inversion-free estimating equations approach for Gaussian Process models," *J. Comput. Graph. Stat.*, 26(1), 98–107.

Bandyopadhyay, S., and Lahiri, S. N. (2009), "Asymptotic properties of discrete Fourier transforms for spatial data," *Sankhya Ser. A.*, pp. 221–259.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008), "Gaussian predictive process models for large spatial data sets," *J. R. Stat. Soc., Ser. B*, 70(4), 825–848.

Bevilacqua, M., and Gaetan, C. (2015), "Comparing composite likelihood methods based on pairs for spatial Gaussian random fields," *Stat. Comput.*, 25(5), 877–892.

Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton, N.J.: Princeton Univ. Press.

Dahlhaus, R. (1983), "Spectral analysis with tapered data," *J. Time Ser. Anal.*, 4(3), 163–175.

Dahlhaus, R., and Künsch, H. (1987), "Edge effects and efficient parameter estimation for stationary random fields," *Biometrika*, 74(4), 877–882.

Deb, S., Pourahmadi, M., and Wu, W. B. (2017), "An asymptotic theory for spectral analysis of random fields," *Electron. J. Stat.*, 11(2), 4297–4322.

Eggers, G. L. (2013), A regionalized maximum-likelihood estimation of the spatial structure of Venusian Topography,, A. B. Thesis, Princeton University.

Fernández-Casal, R., and Crujeiras, R. M. (2010), "Spatial dependence estimation using FFT of biased covariances," *J. Stat. Plan. Inf.*, 140(9), 2653–2668.

Fuentes, M. (2007), "Approximate likelihood for large irregularly spaced spatial data," *J. Am. Stat. Assoc.*, 102(477), 321–331.

GEBCO Bathymetric Compilation Group (2019), The GEBCO_2019 grid—A continuous terrain model of the global oceans and land,, Technical report, British Oceanographic Data Centre, National Oceanography Centre, NERC.

Giraitis, L., and Taqqu, M. S. (1999), "Whittle estimator for finite-variance non-Gaussian time series with long memory," *Annals of Statistics*, 27, 178–203.

Grenander, U., and Szegö, G. (1958), *Toeplitz Forms and Their Applications*, Berkeley, Calif.: Univ. Calif. Press.

Guillaumin, A. P., Sykulski, A. M., Olhede, S. C., Early, J. J., and Lilly, J. M. (2017), "Analysis of non-stationary modulated time series with applications to oceanographic surface flow measurements," *J. Time Ser. Anal.*, 38(5), 668–710.

Guinness, J. (2019), "Spectral density estimation for random fields via periodic embeddings," *Biometrika*, 106, 267–286.

Guinness, J., and Fuentes, M. (2017), "Circulant embedding of approximate covariances for inference from Gaussian data on large lattices," *J. Comput. Graph. Stat.*, 26(1), 88–97.

Guyon, X. (1982), "Parameter estimation for a stationary process on a d-dimensional lattice," *Biometrika*, 69(1), 95–105.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., and Zammit-Mangion, A. (2019), "A case study competition among methods for analyzing large spatial data," *J. Agr. Biol. Envir. St.*, 24(3), 398–425.

Heyde, C. C. (1997), *Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation*, New York: Springer.

Horn, R. A., and Johnson, C. R. (1985), *Matrix analysis*, Cambridge, UK: Cambridge Univ. Press.

Jesus, J., and Chandler, R. E. (2017), "Inference with the Whittle likelihood: A tractable approach using estimating functions," *J. Time Ser. Anal.*, 38(2), 204–224.

Katzfuss, M. (2017), "A multi-resolution approximation for massive spatial datasets," *J. Am. Stat. Assoc.*, 112(517), 201–214.

Kaufman, C. G., Schervish, M. J., and Nychka, D. W. (2008), "Covariance tapering for likelihood-based estimation in large spatial data sets," *J. Am. Stat. Assoc.*, 103(484), 1545–1555.

Kazeev, V. A., Khoromskij, B. N., and Tyrtyshnikov, E. E. (2013), "Multilevel Toeplitz matrices generated by tensor-structured vectors and convolution with logarithmic complexity," *SIAM J. Sci. Comput.*, 35(3), A1511–A1536.

Kent, J. T., and Mardia, K. V. (1996), "Spectral and circulant approximations to the likelihood for stationary Gaussian random fields," *J. Stat. Plan. Inf.*, 50(3), 379–394.

Koopmans, L. H. (1995), *The Spectral Analysis of Time Series*, 2 edn, San Diego, Calif.: Academic Press.

Lee, D., and Mitchell, R. (2013), "Locally adaptive spatial smoothing using conditional auto-regressive models," *J. R. Stat. Soc., Ser. C*, 62(4), 593–608.

Mardia, K. V., and Marshall, R. J. (1984), "Maximum likelihood estimation of models for residual covariance in spatial regression," *Biometrika*, 71(1), 135–146.

Matsuda, Y., and Yajima, Y. (2009), "Fourier analysis of irregularly spaced data on $\mathbb{R}^d$," *J. R. Stat. Soc., Ser. B*, 71(1), 191–217.

Rao, S. S. (2018), "Statistical inference for spatial statistics defined in the Fourier domain," *Ann. Stat.*, 46(2), 469–499.

Rappaport, N. J., Konopliv, A. S., Kucinskas, A. B., and Ford, P. G. (1999), "An improved 360 degree and order model of Venus topography," *Icarus*, 139, 19–31.

Robinson, P. M., and Sanz, J. V. (2006), "Modified Whittle estimation of multilateral models on a lattice," *J. Multivariate Anal.*, 97(5), 1090–1120.

Sang, H., and Huang, J. Z. (2012), "A full scale approximation of covariance functions for large spatial data sets," *J. R. Stat. Soc., Ser. B*, 74(1), 111–132.

Shaby, B., and Ruppert, D. (2012), "Tapered covariance: Bayesian estimation and asymptotics," *J. Comput. Graph. Stat.*, 21(2), 433–452.

Simons, F. J., and Olhede, S. C. (2013), "Maximum-likelihood estimation of lithospheric flexural rigidity, initial-loading fraction and load correlation, under isotropy," *Geophys. J. Int.*, 193(3), 1300–1342.

Stein, M. L., Chi, Z., and Welty, L. J. (2004), "Approximating likelihoods for large spatial data sets," *J. R. Stat. Soc., Ser. B*, 66(2), 275–296.

Stroud, J. R., Stein, M. L., and Lysen, S. (2017), "Bayesian and maximum likelihood estimation for Gaussian processes on an incomplete lattice," *J. Comput. Graph. Stat.*, 26(1), 108–120.

Sykulski, A. M., Olhede, S. C., Guillaumin, A. P., Lilly, J. M., and Early, J. J. (2019), "The debiased Whittle likelihood," *Biometrika*, 106(2), 251–266.

Tyrtyshnikov, E. E., and Zamarashkin, N. L. (1998), "Spectra of multilevel Toeplitz matrices: Advanced theory via simple matrix relationships," *Lin. Alg. & Appl.*, 270(1-3), 15–27.

Varin, C., Reid, N., and Firth, D. (2011), "An overview of composite likelihood methods," *Stat. Sinica.*, 21, 5–42.

Zhang, H. (2004), "Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics," *J. Am. Stat. Assoc.*, 99(465), 250–261.

Zhang, H., and Zimmerman, D. L. (2005), "Towards reconciling two asymptotic frameworks in spatial statistics," *Biometrika*, 92(4), 921–936.

Proofs of lemmata, propositions and theorems

Proof of lemma 2

Proof. Let $\mathbf{k} \in \prod_{i=0}^{d-1} \{0, \ldots, n_k - 1\}$. We remind the reader that for $\mathbf{u} \in \mathbb{Z}^d$, $\bar{c}_{\mathbf{n}}(\mathbf{u}) = c_{g,\mathbf{n}_k}(\mathbf{u})c_X(\mathbf{u})$, where,

$$c_{g,\mathbf{n}_k}(\mathbf{u}) = \frac{\sum_{\mathbf{s}\in\mathbb{Z}^d} g_{\mathbf{s}} g_{\mathbf{s}+\mathbf{u}}}{\sum_{\mathbf{s}\in\mathbb{Z}^d} g_{\mathbf{s}}^2}.$$

Using the fact that for any $\mathbf{q} \in \{0, 1\}^d$,

$$\bar{c}_{\mathbf{n}}(\mathbf{u} - \mathbf{q} \circ \mathbf{n}) \exp\left(-i\sum_{j=0}^{d-1} \frac{2k_j\pi}{n_j}(u_j - q_j n_j)\right) = \bar{c}_{\mathbf{n}}(\mathbf{u} - \mathbf{q} \circ \mathbf{n}) \exp\left(-i\sum_{j=0}^{d-1} \frac{2k_j\pi}{n_j}u_j\right),$$

and since $\bar{c}_{\mathbf{n}}(\mathbf{u}-\mathbf{q}\circ\mathbf{n})$ is zero if any component of $\mathbf{u}$ is zero and the corresponding component of $\mathbf{q}$ is one (due to the definition of $c_{g,\mathbf{n}_k}$), we obtain the proposed formula. Indeed, any $\mathbf{u} \in \prod_{i=0}^{d-1} \{-(n_k - 1), \ldots, n_k - 1\}$ that contribute to the LHS of the proposed formula can be written as $\mathbf{u} = \mathbf{u}^+ - \mathbf{q}\circ\mathbf{n}$ for some unique $\mathbf{u}^+ \in \prod_{i=0}^{d-1} \{0, \ldots, n_k - 1\}$. The extra terms in the RHS of the proposed formula take value zero according to the previous argument. $\square$

Proof of Lemma 5

Proof.

$$\sum_{\mathbf{u}\in\mathbb{Z}} c_g(\mathbf{u}) = \frac{1}{\sum_{\mathbf{s}\in\mathbb{Z}} g_{\mathbf{s}}^2} \sum_{\mathbf{u}\in\mathbb{Z}} \sum_{\mathbf{s}\in\mathbb{Z}} g_{\mathbf{s}} g_{\mathbf{s}+\mathbf{u}} = \frac{1}{\sum_{\mathbf{s}\in\mathbb{Z}} g_{\mathbf{s}}^2} \sum_{\mathbf{s}\in\mathbb{Z}} g_{\mathbf{s}} \sum_{\mathbf{u}\in\mathbb{Z}} g_{\mathbf{s}+\mathbf{u}}$$

$$= \frac{\left(\sum_{\mathbf{s}\in\mathbb{Z}} g_{\mathbf{s}}\right)^2}{\sum_{\mathbf{s}\in\mathbb{Z}} g_{\mathbf{s}}^2} \geq \frac{\left(\sum_{\mathbf{s}\in\mathbb{Z}} g_{\mathbf{s}}\right)^2}{\sum_{\mathbf{s}\in\mathbb{Z}} g_{\mathbf{s}}} = \sum_{\mathbf{s}\in\mathbb{Z}} g_{\mathbf{s}},$$

where we have used the fact that $g_{\mathbf{s}}$ is assumed to be upper-bounded by one. $\square$

Proof of Lemma 6

Proof. This comes as a consequence of the fact that for a sequence of full rectangular grids that grow unbounded in all directions, for any $\mathbf{u} \in \mathbb{Z}^d$ we have $c_{g,n_k}(\mathbf{u}) \to 1$ as k goes to infinity, see equation (15) in the main body. We then conclude by application of Parseval's equality. □

Proof of Lemma 7

Proof. The argument is very similar to that of Lemma 6, with the difference that for any $\mathbf{u} \in \mathbb{Z}^d$ we have that $c_{g,n_k}(\mathbf{u})$ converges to a positive constant (which might be strictly smaller than one) as k goes to infinity. □

Proof of Theorem 1

Proof. We will show in Lemma 14 that $l_{\mathbf{n}_k}(\cdot)$ converges uniformly to $\widetilde{l}_{\mathbf{n}_k}(\cdot)$ in probability, i.e., their difference converges uniformly to the zero function in probability. Hence the difference $l_{\mathbf{n}_k}(\hat{\boldsymbol{\theta}}_k) - \widetilde{l}_{\mathbf{n}_k}(\hat{\boldsymbol{\theta}}_k)$ converges to zero in probability. Additionally, $l_{\mathbf{n}_k}(\hat{\boldsymbol{\theta}}_k) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\theta})$ converges to zero in probability. Indeed, by definition, the parameter vector $\hat{\boldsymbol{\theta}}_k$ minimizes the function $l_{\mathbf{n}_k}(\cdot)$ over the parameter set Θ, and according to Lemma 9, the parameter vector $\boldsymbol{\theta}$ minimizes the function $\widetilde{l}_{\mathbf{n}_k}(\cdot)$. We therefore have, by the triangle inequality,

$$
\left| \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\theta}) - \widetilde{l}_{\mathbf{n}_k}(\hat{\boldsymbol{\theta}}_k) \right| \leq \left| l_{\mathbf{n}_k}(\hat{\boldsymbol{\theta}}_k) - \widetilde{l}_{\mathbf{n}_k}(\hat{\boldsymbol{\theta}}_k) \right| + \left| l_{\mathbf{n}_k}(\hat{\boldsymbol{\theta}}_k) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\theta}) \right|,
$$

which converges to zero in probability. Making use of Lemma 13 we conclude that $\hat{\boldsymbol{\theta}}_k$ converges in probability to $\boldsymbol{\theta}$. □

Proof of Proposition 1

Proof. Let $a_{\max} > 0$ be a finite constant such that $|a_{\mathbf{n}}(\boldsymbol{\omega})| \leq a_{\max}, \forall \boldsymbol{\omega} \in \mathcal{T}^2, \forall \mathbf{n} \in \mathbb{N}^d$. We first make the observation that the sum of the periodogram values at the Fourier frequencies is the squared L_2 norm of the sample, up to some multiplicative constant, since the Discrete Fourier Transform is orthonormal, i.e.

$$
\sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}}} I_{\mathbf{n}}(\boldsymbol{\omega}) = \frac{|\mathbf{n}|}{(2\pi)^d \sum_{\mathbf{s} \in \mathbb{Z}^d} g_{\mathbf{s}}^2} \sum_{\mathbf{s} \in \mathbb{Z}^d} g_{\mathbf{s}}^2 X_{\mathbf{s}}^2.
$$

Therefore,

$$\text{var}\left\{|\mathbf{n}|^{-1}\sum_{\boldsymbol{\omega}\in\Omega_{\mathbf{n}}}a_{\mathbf{n}}(\boldsymbol{\omega})I_{\mathbf{n}}(\boldsymbol{\omega})\right\}\leq a_{\max}^2\text{var}\left\{|\mathbf{n}|^{-1}\sum_{\boldsymbol{\omega}\in\Omega_{\mathbf{n}}}I_{\mathbf{n}}(\boldsymbol{\omega})\right\}$$

$$=\frac{a_{\max}^2}{(2\pi)^{2d}\left(\sum g_{\mathbf{s}}^2\right)^2}\text{var}\left\{\sum_{\mathbf{s}\in\mathbb{Z}^d}g_{\mathbf{s}}^2X_{\mathbf{s}}^2\right\}. \qquad (34)$$

Note that the first inequality is valid since the covariance of the periodogram at two Fourier frequencies $\boldsymbol{\omega},\boldsymbol{\omega}'$ is non-negative for a Gaussian process. Indeed, letting

$$J(\boldsymbol{\omega})=\frac{(2\pi)^{-d/2}}{\sqrt{\sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}g_{\mathbf{s}}^2}}\sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}g_{\mathbf{s}}X_{\mathbf{s}}\exp(-i\boldsymbol{\omega}\cdot\mathbf{s}),$$

we have, by Isserlis' theorem,

$$\begin{aligned}
\text{cov}\left\{I(\boldsymbol{\omega}),I(\boldsymbol{\omega}')\right\}&=\text{E}\left\{J(\boldsymbol{\omega})J^*(\boldsymbol{\omega})J(\boldsymbol{\omega}')J^*(\boldsymbol{\omega}')\right\}-\text{E}\left\{I(\boldsymbol{\omega})I(\boldsymbol{\omega}')\right\}\\
&=\text{E}\left\{J(\boldsymbol{\omega})J(\boldsymbol{\omega}')\right\}\text{E}\left\{J^*(\boldsymbol{\omega})J^*(\boldsymbol{\omega}')\right\}+\text{E}\left\{J(\boldsymbol{\omega})J^*(\boldsymbol{\omega}')\right\}\text{E}\left\{J^*(\boldsymbol{\omega})J(\boldsymbol{\omega}')\right\}\\
&=\left|\text{E}\left\{J(\boldsymbol{\omega})J(\boldsymbol{\omega}')\right\}\right|^2+\left|\text{E}\left\{J^*(\boldsymbol{\omega})J(\boldsymbol{\omega}')\right\}\right|^2,
\end{aligned}$$

which is non-negative as the sum of two squares. We study the term $\text{var}\left\{\sum_{\mathbf{s}\in\mathbb{Z}^d}g_{\mathbf{s}}^2X_{\mathbf{s}}^2\right\}$. We have, again using Isserlis' theorem for Gaussian random variables,

$$\begin{aligned}
\text{var}\left\{\sum_{\mathbf{s}\in\mathbb{Z}^d}g_{\mathbf{s}}^2X_{\mathbf{s}}^2\right\}&=\text{E}\left(\sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}g_{\mathbf{s}}^2X_{\mathbf{s}}^2\right)^2-\left(\text{E}\sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}g_{\mathbf{s}}^2X_{\mathbf{s}}^2\right)^2\\
&=\sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}\sum_{\mathbf{s}'\in\mathcal{J}_{\mathbf{n}}}\text{E}\left\{g_{\mathbf{s}}^2g_{\mathbf{s}'}^2X_{\mathbf{s}}^2X_{\mathbf{s}'}^2\right\}-\text{E}\left\{g_{\mathbf{s}}^2X_{\mathbf{s}}^2\right\}\text{E}\left\{g_{\mathbf{s}'}^2X_{\mathbf{s}'}^2\right\}\\
&=2\sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}\sum_{\mathbf{s}'\in\mathcal{J}_{\mathbf{n}}}g_{\mathbf{s}}^2g_{\mathbf{s}'}^2\left(\text{E}\left\{X_{\mathbf{s}}X_{\mathbf{s}'}\right\}\right)^2 \qquad (35)
\end{aligned}$$

We now obtain, combining equations (34) and (35),

$$\mathrm{var}\left\{|\mathbf{n}|^{-1}\sum_{\boldsymbol{\omega}\in\Omega_{\mathbf{n}}}a_{\mathbf{n}}(\boldsymbol{\omega})I_{\mathbf{n}}(\boldsymbol{\omega})\right\}\leq\frac{2a_{\max}^2}{(2\pi)^{2d}\left(\sum g_{\mathbf{s}}^2\right)^2}\sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}\sum_{\mathbf{s}'\in\mathcal{J}_{\mathbf{n}}}g_{\mathbf{s}}^2g_{\mathbf{s}'}^2\left(\mathrm{E}\left\{X_{\mathbf{s}}X_{\mathbf{s}'}\right\}\right)^2$$

$$\leq\frac{2a_{\max}^2}{(2\pi)^{2d}\left(\sum g_{\mathbf{s}}^2\right)^2}\sum_{\mathbf{u}\in\mathbb{Z}^d}c_X(\mathbf{u})^2\sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}g_{\mathbf{s}}^2g_{\mathbf{s}+\mathbf{u}}^2$$

$$\leq\frac{2a_{\max}^2}{(2\pi)^{2d}\left(\sum g_{\mathbf{s}}^2\right)^2}\sum_{\mathbf{u}\in\mathbb{Z}^d}c_X(\mathbf{u})^2g_{\max}^2\sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}g_{\mathbf{s}}g_{\mathbf{s}+\mathbf{u}}$$

$$\leq\frac{2a_{\max}^2}{(2\pi)^{2d}\sum g_{\mathbf{s}}^2}\sum_{\mathbf{u}\in\mathbb{Z}^d}c_X(\mathbf{u})^2c_g(\mathbf{u}),$$

where we have made use of the assumption that $0\leq g_{\mathbf{s}}\leq 1, \forall \mathbf{s}\in\mathbb{Z}^d$. Therefore, we obtain the stated result, i.e.,

$$\mathrm{var}\left\{|\mathbf{n}_k|^{-1}\sum_{\boldsymbol{\omega}\in\Omega_{\mathbf{n}_k}}a_k(\boldsymbol{\omega})I_{\mathbf{n}_k}(\boldsymbol{\omega})\right\}=\mathcal{O}\left\{\frac{\sum_{\mathbf{u}\in\mathbb{Z}^d}c_X(\mathbf{u})^2c_{g,k}(\mathbf{u})}{\sum g_{\mathbf{s}}^2}\right\},$$

where the big O is with respect to k going to infinity. $\qquad\square$

Proof of Lemma 9

Proof. The difference between the expected likelihood function at the true parameter vector and any parameter vector $\boldsymbol{\gamma}\in\Theta$ takes the form

$$\widetilde{l}_{\mathbf{n}}(\boldsymbol{\gamma})-\widetilde{l}_{\mathbf{n}}(\boldsymbol{\theta})=|\mathbf{n}|^{-1}\sum_{\boldsymbol{\omega}\in\Omega_{\mathbf{n}}}\phi\left(\frac{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma})}\right),$$

with $\phi:x\mapsto x-\log x-1$. This function is non-negative and attains it minimum uniquely at $x=1$. $\qquad\square$

Proof of Lemma 10

Proof. By combining equations (4) and (12) in the main body then the periodogram can be expressed as

$$I_{\mathbf{n}}(\boldsymbol{\omega})=\frac{(2\pi)^{-d}}{\sum g_{\mathbf{s}}^2}\left|\sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}g_{\mathbf{s}}X_{\mathbf{s}}\exp(-i\boldsymbol{\omega}\cdot\mathbf{s})\right|^2,\qquad\boldsymbol{\omega}\in\mathcal{T}^d.$$

Making use of equation (9) of the main body, we therefore have,

$$\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma}) = \int_{\mathcal{T}^d} f_{\delta,X}(\boldsymbol{\omega}-\boldsymbol{\lambda};\boldsymbol{\gamma})\mathcal{F}_{\mathbf{n}_k}(\boldsymbol{\lambda})d\boldsymbol{\lambda}.$$

Also,

$$\begin{aligned}
\int_{\mathcal{T}^d} \mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega})d\boldsymbol{\omega} &= \frac{(2\pi)^{-d}}{\sum g_{\mathbf{s}}^2} \int_{\mathcal{T}^d} \left| \sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}} g_{\mathbf{s}}\exp(i\boldsymbol{\omega}\cdot\mathbf{s}) \right|^2 d\boldsymbol{\omega} \\
&= \frac{(2\pi)^{-d}}{\sum g_{\mathbf{s}}^2} \int_{\mathcal{T}^d} \sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}\sum_{\mathbf{s}'\in\mathcal{J}_{\mathbf{n}}} g_{\mathbf{s}}g_{\mathbf{s}'}\exp\{i\boldsymbol{\omega}\cdot(\mathbf{s}'-\mathbf{s})\}d\boldsymbol{\omega} \\
&= \frac{(2\pi)^{-d}}{\sum g_{\mathbf{s}}^2} \sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}\sum_{\mathbf{s}'\in\mathcal{J}_{\mathbf{n}}} \int_{\mathcal{T}^d} g_{\mathbf{s}}g_{\mathbf{s}'}\exp\{i\boldsymbol{\omega}\cdot(\mathbf{s}'-\mathbf{s})\}d\boldsymbol{\omega} \\
&= \frac{1}{\sum g_{\mathbf{s}}^2} \sum_{\mathbf{s}\in\mathcal{J}_{\mathbf{n}}}\sum_{\mathbf{s}'\in\mathcal{J}_{\mathbf{n}}} g_{\mathbf{s}}g_{\mathbf{s}'}\delta_{\mathbf{s},\mathbf{s}'} \\
&= 1,
\end{aligned}$$

which is a direct adaptation of a standard result for the Féjer kernel. Hence,

$$\begin{aligned}
\left|\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma})\right| &\leq \int_{\mathcal{T}^d} \left|f_{\delta,X}(\boldsymbol{\omega}-\boldsymbol{\lambda};\boldsymbol{\gamma})\mathcal{F}_{\mathbf{n}_k}(\boldsymbol{\lambda})\right|d\boldsymbol{\lambda} \\
&\leq f_{\delta,\max} \int_{\mathcal{T}^d} \left|\mathcal{F}_{\mathbf{n}_k}(\boldsymbol{\lambda})\right|d\boldsymbol{\lambda} \\
&\leq f_{\delta,\max}.
\end{aligned}$$

Similarly, we obtain the other inequality, i.e. $\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma}) \geq f_{\delta,\min}$, which concludes the proof. $\qquad\square$

Proof of Lemma 11

Proof. We first observe, given equation (23) of the main body, that

$$\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\theta}) = |\mathbf{n}_k|^{-1} \sum_{\boldsymbol{\omega}\in\Omega_{\mathbf{n}_k}} \left\{ \frac{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma})} - \log\frac{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma})} - 1 \right\}.$$

As before, denoting $\phi : x \mapsto x - \log x - 1$, $x > 0$, and $g_{\mathbf{n}}(\boldsymbol{\omega})$ the piece-wise continuous function that maps any frequency of $\mathcal{T}^d$ to the closest smaller Fourier frequency corresponding to the grid $\mathcal{J}_{\mathbf{n}}$, we have

$$\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\theta}) = (2\pi)^{-d} \int_{\mathcal{T}^d} \phi\left(\frac{\overline{I}_{\mathbf{n}_k}(g(\boldsymbol{\omega});\boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(g(\boldsymbol{\omega});\boldsymbol{\gamma})}\right)d\boldsymbol{\omega}.$$

A Taylor expansion of $\phi(\cdot)$ around 1 gives, with $\psi(x) = (x-1)^2$,

$$\phi(x) = \psi(x)(1 + \epsilon(x)),$$

where $\epsilon(x) \to 0$ as $x \to 1$. Therefore for any $\delta > 0$ there exists $\mu > 0$ such that for all x such that $|x - 1| \le \mu$, $|\epsilon(x)| < \delta$. Now let, for all $k \in \mathbb{N}$,

$$\Pi_k = \left\{ \boldsymbol{\omega} \in \mathcal{T}^d : \left| \frac{\overline{I}_{\mathbf{n}_k}(g(\boldsymbol{\omega}); \boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(g(\boldsymbol{\omega}); \boldsymbol{\gamma})} - 1 \right| \le \mu \right\}.$$

We distinguish two cases:

1. If for some $\delta > 0$, the Lebesgue measure of Π_k does not converge to $(2\pi)^d$, equation (25) of the main body holds.

2. Otherwise, if for any $\delta > 0$ the Lebesgue measure of Π_k does converge to $(2\pi)^d$, we then have

$$\left| \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\theta}) \right| = \int_{\Pi_k \cup \Pi_k^C} \psi \left(\frac{\overline{I}_{\mathbf{n}_k}(g(\boldsymbol{\omega}); \boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(g(\boldsymbol{\omega}); \boldsymbol{\gamma})} \right) \left\{ 1 + \epsilon \left(\frac{\overline{I}_{\mathbf{n}_k}(g(\boldsymbol{\omega}); \boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(g(\boldsymbol{\omega}); \boldsymbol{\gamma})} \right) \right\} d\boldsymbol{\omega},$$

where Π_k^C denotes the complementary of Π_k as a subset of $\mathcal{T}^d$ and where the function $\epsilon(\cdot)$ was defined in equation (7). Denoting $h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma}) = \frac{\overline{I}_{\mathbf{n}_k}(g(\boldsymbol{\omega}); \boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(g(\boldsymbol{\omega}); \boldsymbol{\gamma})}$ (note that this quantity also depends on k),

$$\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\theta}) = \int_{\mathcal{T}^d} \psi\left(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma}) \right) d\boldsymbol{\omega}$$
$$+ \int_{\Pi_k} \psi\left(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma}) \right) \epsilon\left(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma}) \right) d\boldsymbol{\omega}$$
$$+ \int_{\Pi_k^C} \psi(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma}) \epsilon(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma})) d\boldsymbol{\omega}.$$

We shall now show that the two last terms of the right-hand side of this equation are asymptotically vanishing, so that we can limit our study to the first term, which will turn out to take a simple form in relation to our definition of significant correlation contribution (SCC) in the main body. Given the definition of Π_k we have,

$$\left| \int_{\Pi_k} \psi\left(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma}) \right) \epsilon\left(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma}) \right) d\boldsymbol{\omega} \right| \le \delta \int_{\Pi_k} \psi\left(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma}) \right) d\boldsymbol{\omega} \le \delta \int_{\mathcal{T}^d} \psi\left(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma}) \right) d\boldsymbol{\omega},$$

where the two inequalities come from the fact that the function $\psi(\cdot)$ is non-negative. We also have

$$\left| \int_{\Pi_k^C} \psi(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma}) \epsilon(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma})) d\boldsymbol{\omega} \right| = o(1),$$

since the integrand is upper-bounded given Assumption 1.2 and since the measure of the set Π_k^C goes to zero. Hence we obtain, by the triangle inequality,

$$\left|\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\theta})\right| \geq \left(\int_{T^d} \psi\left(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma})\right) d\boldsymbol{\omega}\right)(1 - \delta) + o(1).$$

We now study the term $(2\pi)^{-d} \int_{\mathcal{T}^d} \psi\left(h(\boldsymbol{\omega}; \boldsymbol{\theta}, \boldsymbol{\gamma})\right) d\boldsymbol{\omega} = |\mathbf{n}_k|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} \left\{\frac{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\gamma})} - 1\right\}^2$. We observe that

$$
\begin{aligned}
|\mathbf{n}_k|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} \left\{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta}) - \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\gamma})\right\}^2 &= |\mathbf{n}_k|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\gamma})^2 \left\{\frac{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\gamma})} - 1\right\}^2 \\
&\leq |\mathbf{n}_k|^{-1} f_{\max,\delta}^2 \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} \left\{\frac{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\gamma})} - 1\right\}^2.
\end{aligned}
$$

Additionally, by Parseval's equality,

$$
\begin{aligned}
|\mathbf{n}_k|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} \left\{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta}) - \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\gamma})\right\}^2 &= \sum_{\mathbf{u} \in \mathbb{Z}^d} \left\{\overline{c}_{\mathbf{n}_k}(\mathbf{u}; \boldsymbol{\theta}) - \overline{c}_{\mathbf{n}_k}(\mathbf{u}; \boldsymbol{\gamma})\right\}^2 \\
&= \sum_{\mathbf{u} \in \mathbb{Z}^d} c_{g,\mathbf{n}_k}(\mathbf{u})^2 \left\{c_X(\mathbf{u}; \boldsymbol{\theta}) - c_X(\mathbf{u}; \boldsymbol{\gamma})\right\}^2 \\
&\geq \frac{1}{2} \underline{\lim}_{k \to \infty} S_k(\boldsymbol{\theta}, \boldsymbol{\gamma}),
\end{aligned}
$$

where the last inequality holds for k sufficiently large, given the SCC assumption, see Definition 1. Therefore we obtain for k sufficiently large,

$$\left|\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\theta})\right| \geq \frac{1}{2 f_{\max,\delta}^2}(1 - \delta) \underline{\lim}_{k \to \infty} S_k(\boldsymbol{\theta}, \boldsymbol{\gamma}) + o(1).$$

Choosing $\delta = 1/2$, we obtain the inequality stated in equation (25) of the main body. This concludes the proof.

$\square$

Proof of Lemma 12

Proof. First we observe that for any fixed $\boldsymbol{\omega} \in \mathcal{T}^d$, $\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\gamma}_k)$ converges to $\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\gamma})$ as k goes to infinity. This comes from Assumption 1.2, where we have assumed an upper-bound

on the derivative of the spectral density with respect to the parameter vector. In that case,

$$
\begin{aligned}
\left|\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma}_k) - \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma})\right| &\leq \left|(2\pi)^{-d}\int_{\mathcal{T}^d}\left\{f_{X,\boldsymbol{\delta}}(\boldsymbol{\omega}-\boldsymbol{\omega}';\boldsymbol{\gamma}_k) - f_{X,\boldsymbol{\delta}}(\boldsymbol{\omega}-\boldsymbol{\omega}';\boldsymbol{\gamma})\right\}\mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega}')d\boldsymbol{\omega}'\right| \\
&\leq (2\pi)^{-d}\int_{\mathcal{T}^d}\left|f_{X,\boldsymbol{\delta}}(\boldsymbol{\omega}-\boldsymbol{\omega}';\boldsymbol{\gamma}_k) - f_{X,\boldsymbol{\delta}}(\boldsymbol{\omega}-\boldsymbol{\omega}';\boldsymbol{\gamma})\right|\mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega}')d\boldsymbol{\omega}' \\
&\leq (2\pi)^{-d}\int_{\mathcal{T}^d}M_{\partial_\theta}\|\boldsymbol{\gamma}_k - \boldsymbol{\gamma}\|_2\mathcal{F}_{\mathbf{n}}(\boldsymbol{\omega}')d\boldsymbol{\omega}' \\
&\leq M_{\partial_\theta}\|\boldsymbol{\gamma}_k - \boldsymbol{\gamma}\|_2
\end{aligned}
$$

which converges to zero as $\|\boldsymbol{\gamma}_k - \boldsymbol{\gamma}\|_2$ converges to zero by assumption.

Now using equation (23), we can apply the Dominated Convergence Theorem to $(\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}_k) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}))_{k\in\mathbb{N}}$, using the bounds established in Lemma 10, and the $\boldsymbol{\omega}$-pointwise convergence of $\left|\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma}_k) - \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma}_k)\right|$ to zero. Hence $(\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}_k) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}))_{k\in\mathbb{N}}$ converges to zero, which concludes the proof. □

Proof of Lemma 13

Proof. Assume, with the intent to reach a contradiction, that $(\boldsymbol{\gamma}_k)$ does not converge to $\boldsymbol{\theta}$. By compactness of Θ, there exists $\boldsymbol{\gamma}\in\Theta$ distinct from $\boldsymbol{\theta}$ and $(\boldsymbol{\gamma}_{j_k})$ a subsequence of $(\boldsymbol{\gamma}_k)$ such that $\boldsymbol{\gamma}_{j_k}$ converges to $\boldsymbol{\gamma}$. We then have, using the inverse triangle inequality,

$$
|\widetilde{l}_{\mathbf{n}_{j_k}}(\boldsymbol{\gamma}_{j_k}) - \widetilde{l}_{\mathbf{n}_{j_k}}(\boldsymbol{\theta})| \geq \left|\widetilde{l}_{\mathbf{n}_{j_k}}(\boldsymbol{\gamma}) - \widetilde{l}_{\mathbf{n}_{j_k}}(\boldsymbol{\theta})\right| - \left|\widetilde{l}_{\mathbf{n}_{j_k}}(\boldsymbol{\gamma}_{j_k}) - \widetilde{l}_{\mathbf{n}_{j_k}}(\boldsymbol{\gamma})\right|.
$$

The second term on the right-hand side of the above equation converges to zero according to Lemma 12 whereas the first term is asymptotically lower bounded according to Lemma 11. Therefore the quantity $|\widetilde{l}_{\mathbf{n}_{j_k}}(\boldsymbol{\gamma}_{j_k}) - \widetilde{l}_{\mathbf{n}_{j_k}}(\boldsymbol{\theta})|$ is asymptotically lower bounded, which contradicts the initial assumption that $\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}_k) - \widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\theta})$ converges to zero. This concludes the proof, by obtaining a contradiction. □

Proof of Lemma 14

Proof. We have

$$
\begin{aligned}
\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - l_{\mathbf{n}_k}(\boldsymbol{\gamma}) &= |\mathbf{n}_k|^{-1}\sum_{\boldsymbol{\omega}\in\Omega_{\mathbf{n}}}\left\{\log\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma}) + \frac{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\theta})}{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma})} - \log\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma}) - \frac{I_{\mathbf{n}_k}(\boldsymbol{\omega})}{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma})}\right\} \\
&= |\mathbf{n}_k|^{-1}\sum_{\boldsymbol{\omega}\in\Omega_{\mathbf{n}}}\frac{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\theta}) - I_{\mathbf{n}_k}(\boldsymbol{\omega})}{\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma})}.
\end{aligned}
$$

We note that

$$
\mathrm{E}\left\{\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - l_{\mathbf{n}_k}(\boldsymbol{\gamma})\right\} = 0.
$$

Given that the quantity $\overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega};\boldsymbol{\gamma})^{-1}$ is deterministic and upper-bounded independently of $\boldsymbol{\gamma}$ by $f_{\min,\delta}^{-1}$, we can use Proposition 1 to conclude that

$$\mathrm{var}\left\{\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - l_{\mathbf{n}_k}(\boldsymbol{\gamma})\right\} = \mathcal{O}\left\{\frac{\sum_{\mathbf{u}\in\mathbb{Z}^d} c_X(\mathbf{u})^2 c_g(\mathbf{u})}{\sum g_{\mathbf{s}}^2}\right\},$$

where the big $\mathcal{O}$ does not depend on $\boldsymbol{\gamma}$. Thus using Chebychev's inequality

$$\widetilde{l}_{\mathbf{n}_k}(\boldsymbol{\gamma}) - l_{\mathbf{n}_k}(\boldsymbol{\gamma}) = \mathcal{O}_P\left\{\left(\frac{\sum_{\mathbf{u}\in\mathbb{Z}^d} c_X(\mathbf{u})^2 c_g(\mathbf{u})}{\sum g_{\mathbf{s}}^2}\right)^{1/2}\right\}$$

This concludes the proof given the SCC assumption.

$\square$

Proof of Lemma 15

Proof. The proof is adapted from the one-dimensional case, see Guillaumin et al. (2017) and Sykulski et al. (2019). We first define the following isomorphism from $\prod_{i=1}^{d}\{1,\cdots,n_i\}$ to $\{1,\cdots,|\mathbf{n}|\}$, that will be used for a change of variable:

$$j(j_1,\ldots,j_d) = \sum_{k=1}^{d}\left\{(j_k-1)\prod_{j=1}^{k-1}n_j\right\},$$

and $j_1(j),\ldots,j_d(j)$ the component functions of its inverse. This isomorphism gives the index in the column vector $\mathbf{X}$ of the observation at location $(j_1,\cdots,j_d)$ on the grid, given our choice of ordering.

Let $\boldsymbol{\alpha}$ be any complex-valued vector of $\mathbb{C}^n$, and denote $\boldsymbol{\alpha}^*$ its Hermitian transpose. We then have, using the above isomorphism for a change of variables,

$$\boldsymbol{\alpha}^* C_{\mathbf{X}}\boldsymbol{\alpha} = \sum_{j,k=1}^{|\mathbf{n}|}\boldsymbol{\alpha}_j^*(C_{\mathbf{X}})_{j,k}\boldsymbol{\alpha}_k$$

$$= \sum_{j_1=0}^{n_1-1}\cdots\sum_{j_d=1}^{n_d-1}\sum_{k_1=0}^{n_1-1}\cdots\sum_{k_d=1}^{n_d-1}\boldsymbol{\alpha}_{j(j_1,\cdots,j_d)}^*(C_{\mathbf{X}})_{j(j_1,\cdots,j_d),k(k_1,\cdots,k_d)}\boldsymbol{\alpha}_{k(k_1,\cdots,k_d)}.$$

Here we use the fact that

$$(C_{\mathbf{X}})_{j(j_1,\cdots,j_d),k(k_1,\cdots,k_d)} = c_{\mathbf{X}}(k_1-j_1,\ldots,k_d-j_d),$$

38

so that

$$\boldsymbol{\alpha}^* C_{\mathbf{X}} \boldsymbol{\alpha} = \sum_{j_1=0}^{n_1-1} \cdots \sum_{j_d=1}^{n_d-1} \sum_{k_1=0}^{n_1-1} \cdots \sum_{k_d=1}^{n_d-1} \boldsymbol{\alpha}^*_{j(j_1,\cdots,j_d)} \boldsymbol{\alpha}_{k(k_1,\cdots,k_d)} \int_{\mathcal{T}^d} f_{X,\delta}(\boldsymbol{\omega}) e^{i((k_1-j_1)\omega_1+\cdots+(k_d-j_d)\omega_d)} d\boldsymbol{\omega}$$

$$= \int_{\mathcal{T}^d} f_{X,\delta}(\boldsymbol{\omega}) \sum_{j_1=0}^{n_1-1} \cdots \sum_{j_d=1}^{n_d-1} \sum_{k_1=0}^{n_1-1} \cdots \sum_{k_d=1}^{n_d-1} \boldsymbol{\alpha}^*_{j(j_1,\cdots,j_d)} \boldsymbol{\alpha}_{k(k_1,\cdots,k_d)} e^{i((k_1-j_1)\omega_1+\cdots+(k_d-j_d)\omega_d)} d\boldsymbol{\omega}$$

$$= \int_{\mathcal{T}^d} f_{X,\delta}(\boldsymbol{\omega}) \left| \sum_{j_1=0}^{n_1-1} \cdots \sum_{j_d=1}^{n_d-1} \boldsymbol{\alpha}_{j(j_1,\cdots,j_d)} e^{i(j_1\omega_1+\cdots+j_d\omega_d)} \right|^2 d\boldsymbol{\omega}$$

$$\leq f_{\delta,\max} \int_{\mathcal{T}^d} \left| \sum_{j_1=0}^{n_1-1} \cdots \sum_{j_d=1}^{n_d-1} \boldsymbol{\alpha}_{j(j_1,\cdots,j_d)} e^{i(j_1\omega_1+\cdots+j_d\omega_d)} \right|^2 d\boldsymbol{\omega}.$$

By Parseval's equality, we obtain,

$$0 \leq \boldsymbol{\alpha}^* C_{\mathbf{X}} \boldsymbol{\alpha} \leq f_{\delta,\max} \|\boldsymbol{\alpha}\|_2^2,$$

where $\|\boldsymbol{\alpha}\|_2$ is the l_2 vector norm of the vector $\boldsymbol{\alpha}$. This concludes the proof of the upper bound. The lower bound can be derived in the same way, which concludes the proof. $\square$

Proof of Proposition 2

Proof. We only treat the scenario where $g_{\mathbf{s}} = 1, \forall \mathbf{s} \in \mathcal{J}_{n_k}$, i.e., we do not consider the situation of missing observations for this proposition. The proof is adapted from Grenander and Szegö (1958, p. 217). Denote

$$L_k = |\mathbf{n}_k|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} w_k(\boldsymbol{\omega}) I_{\mathbf{n}_k}(\boldsymbol{\omega}),$$

as a weighted sum of periodogram values, and $U_{\mathbf{n}_k}$ the multi-dimensional Fourier matrix corresponding to $\mathcal{J}_{\mathbf{n}}$. We have

$$L_k = |\mathbf{n}_k|^{-1} \mathbf{X}^* U_{\mathbf{n}_k}^* \mathrm{diag}(w_k(\boldsymbol{\omega}_0), \ldots, w_k(\boldsymbol{\omega}_{|\mathbf{n}_k|-1})) U_{\mathbf{n}_k} \mathbf{X}.$$

Denoting $W_k = |\mathbf{n}_k|^{-1} U_{\mathbf{n}_k}^* \mathrm{diag}(w_k(\boldsymbol{\omega}_0), \ldots, w_k(\boldsymbol{\omega}_{|\mathbf{n}_k|-1})) U_{\mathbf{n}_k}$, we then have

$$L_k = \mathbf{X}^* W_k \mathbf{X},$$

which we regard as a quadratic form in the vector $\mathbf{X}$. Following Cramér (1946, p. 134), his formula 11.12.2, the characteristic function of the random variable L_k therefore takes the form of

$$\phi_{L_k}(\alpha) = \mathrm{E}\{\exp(i\alpha L_k)\}$$
$$= (2\pi)^{-n/2} |C_X(\boldsymbol{\theta})|^{-1/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-x^*\left(-i\alpha W_k + \frac{1}{2} C_X^{-1}(\boldsymbol{\theta})\right) x\right\} dx_1 \cdots dx_n,$$

39

where for a square matrix A, $|A|$ denotes its determinant. Using a known result (Horn and Johnson, 1985) for complex-valued symmetric matrices, there exists a diagonal matrix D_k and a unitary matrix V_k such that

$$-i\alpha W_k + \frac{1}{2}C_X^{-1}(\boldsymbol{\theta}) = VD_kV^T. \tag{36}$$

By posing the change of variables $y = V^T x$ we obtain,

$$\phi_{L_k}(\alpha) = (2\pi)^{-n/2}\,|C_X(\boldsymbol{\theta})|^{-1/2}\prod_{j=1}^{n}\int_{-\infty}^{\infty}\exp\left\{-y^2 d_{j,k}\right\}dy,$$

where the $d_{j,k}, j = 1, \cdots, n$ are the complex-valued elements of the diagonal matrix D_k from equation (36), and where we remind the reader that $|V| = 1$ since V is unitary. As we recognize integrals of the form $\int_{-\infty}^{\infty}\exp(-y^2)dy$ we obtain,

$$\phi_{L_k}(\alpha) = 2^{-n/2}\,|C_X(\boldsymbol{\theta})|^{-1/2}\left|-i\alpha W_k + \frac{1}{2}C_X(\boldsymbol{\theta})^{-1}\right|^{-1/2}$$

$$= \left|-2i\alpha C_X(\boldsymbol{\theta})W_k + I_{|\mathbf{n}|}\right|^{-1/2}$$

Hence,

$$\log\phi_{L_k}(\alpha) = -\frac{1}{2}\log\left|I_{|\mathbf{n}_k|} - 2i\alpha C_X(\boldsymbol{\theta})W_k\right|.$$

Denoting with $\nu_{1,k},\ldots,\nu_{|\mathbf{n}_k|,k}$ the eigenvalues of $C_X(\boldsymbol{\theta})W_k$, we therefore have

$$\log\phi_{L_k}(\alpha) = -\frac{1}{2}\sum_{j=1}^{|\mathbf{n}_k|}\log\left(1 - 2i\alpha\nu_{j,k}\right).$$

According to Proposition 15 the spectral norm of C_X, the covariance matrix of $\mathbf{X}$, is upper-bounded by $f_{\max,\delta}$. The spectral norm of W_k is clearly upper-bounded by $|\mathbf{n}_k|^{-1}M_W$, as from the definition of W_k its eigenvalues are exactly

$$|\mathbf{n}_k|^{-1}w_k(\boldsymbol{\omega}_0), |\mathbf{n}_k|^{-1}w_k(\boldsymbol{\omega}_1), \cdots, |\mathbf{n}_k|^{-1}w_k(\boldsymbol{\omega}_{|\mathbf{n}_k|-1}).$$

By property of the spectral norm of a product of matrices, we obtain,

$$|\mathbf{n}_k|^{-1}m_W f_{\min,\delta} \le |\nu_{j,k}| \le |\mathbf{n}_k|^{-1}M_W f_{\max,\delta}, \quad \forall j = 1,\cdots,|\mathbf{n}_k|, k \in \mathbb{N}. \tag{37}$$

The variance of L_k is given by

$$\sigma_k^2 = \text{var}\left\{L_k\right\} = 2\sum_{j=1}^{|\mathbf{n}_k|}\nu_{j,k}^2,$$

and therefore satisfies

$$2|\mathbf{n}_k|^{-1}(m_W f_{\min})^2 \le \sigma_k^2 \le 2|\mathbf{n}_k|^{-1}(M_W f_{\max,\delta})^2. \tag{38}$$

40

We also observe that

$$\frac{\nu_{j,k}}{\sigma_k} \to 0, \ (k \to \infty),$$

uniformly, given the bounds determined in equations (37) and (38). Denote $\underline{L}_k$ the standardized quantity $(L_k - \mathrm{E}\{L_k\})/\sigma_k$. After Taylor expansion of the logarithm terms to third order, its characteristic function takes the form of

$$\log \phi_{\underline{L}_k}(\alpha) = -\frac{1}{2} \sum_{j=1}^{|\mathbf{n}_k|} \log \left(1 - \frac{2i\alpha\nu_{j,k}}{\sigma_k}\right) - i\frac{\alpha \sum_{j=1}^{|\mathbf{n}_k|} \nu_{j,k}}{\sigma_k}$$

$$= -\frac{1}{2}\alpha^2 + \sum_{j=1}^{|\mathbf{n}_k|} \left[\frac{4}{3}\left(\frac{i\alpha\nu_{j,k}}{\sigma_k}\right)^3 + o\left\{\left(\frac{i\alpha\nu_{j,k}}{\sigma_k}\right)^3\right\}\right], \tag{39}$$

where the small o is uniform and is denoted ϵ_k in what follows, to make it clear that it does not depend on j. The second term in equation (39) can be shown to become negligible as k goes to infinity, since

$$\left|\sum_{j=1}^{|\mathbf{n}_k|} \left[\frac{4}{3}\left(\frac{i\alpha\nu_{j,k}}{\sigma_k}\right)^3 + o\left\{\left(\frac{i\alpha\nu_{j,k}}{\sigma_k}\right)^3\right\}\right]\right| \le \alpha^3 \sigma_k^{-3}\left(\frac{4}{3} + \epsilon_k\right) \sum_{j=1}^{|\mathbf{n}_k|} |\nu_{j,k}|^3$$

$$\le \alpha^3 \left(\frac{4}{3} + \epsilon_k\right) \frac{|\mathbf{n}_k|^{-2}M_W^3 f_{\max}^3}{|\mathbf{n}_k|^{-3/2}m_W^3 f_{\min}^3}$$

$$= \mathcal{O}(|\mathbf{n}_k|^{-1/2}).$$

We conclude that $\phi_{\underline{L}_k}(\alpha)$ converges to $\exp(-\frac{1}{2}\alpha)$, and therefore L_k is asymptotically standard normally distributed after appropriate normalization. □

Proof of Theorem 2

Proof. Direct calculations show that the gradient of our quasi-likelihood function at the true parameter vector is given by,

$$\nabla_\theta l_{\mathbf{n}_k}(\boldsymbol{\theta}) = |\mathbf{n}_k|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} \bar{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})^{-2} \nabla_\theta \bar{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})\left(\bar{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta}) - I(\boldsymbol{\omega})\right). \tag{40}$$

By expanding this gradient function at the true parameter value, and noting that $\nabla_\theta l_{\mathbf{n}_k}(\boldsymbol{\omega}; \hat{\boldsymbol{\theta}}) = 0$ by definition of $\hat{\boldsymbol{\theta}}$ and given Assumption 2.1, we obtain

$$\nabla_\theta l_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta}) = H(\boldsymbol{\theta}'_k)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_k),$$

where $H(\cdot)$ is the Hessian of $l_{\mathbf{n}_k}(\cdot)$ and $\boldsymbol{\theta}'_k$ is a parameter vector that converges in probability to the true parameter vector, since $\hat{\boldsymbol{\theta}}_k$ is consistent as per Theorem 1. Therefore,

$$\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta} = -H^{-1}(\boldsymbol{\theta}'_k)\nabla_\theta l_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta}). \tag{41}$$

41

We now study the expected Hessian of the likelihood function taken at the true parameter vector, $\mathcal{H}(\boldsymbol{\theta})$. Direct calculations lead to

$$\mathcal{H}(\boldsymbol{\theta}) = |\mathbf{n}_k|^{-1} \sum_{\boldsymbol{\omega} \in \Omega_{\mathbf{n}_k}} \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})^{-2} \nabla_\theta \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta}) \nabla_\theta \overline{I}_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})^T.$$

It can be shown, see Sykulski et al. (2019, p. 17 of their supplementary document) for instance, that in equation (41) the quantity $H(\boldsymbol{\theta}'_k)$ satisfies, if Assumption 2.2 holds,

$$H(\boldsymbol{\theta}'_k) = \mathcal{H}(\boldsymbol{\theta}) + \mathcal{O}_P(r_k) + o_P(1).$$

Hence we have, asymptotically,

$$H^{-1}(\boldsymbol{\theta}'_k) = \mathcal{H}^{-1}(\boldsymbol{\theta}) + o_P(1). \tag{42}$$

Since equation (40) follows the conditions required for Proposition 1 to apply, the gradient at the true parameter vector $\nabla_\theta l_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})$ is itself $\mathcal{O}_P(r_k)$. Further more, Lemma 16 tells us that the minimum eigenvalue of $\mathcal{H}$ is lower-bounded by $S(\boldsymbol{\theta})$, independently of k. We finally obtain the stated result,

$$\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta} = \mathcal{O}_P(r_k).$$

In the case of a sequence of full grids, $|\mathbf{n}_k|^{1/2} |\nabla_\theta l_{\mathbf{n}_k}(\boldsymbol{\omega}; \boldsymbol{\theta})$ is additionally shown to follow a standard normal distribution via Proposition 2, and we conclude to the asymptotic normality of our estimator. $\qquad \square$