

Automatic spectral density estimation for random fields on a lattice via bootstrap

Jose M. Vidal-Sanz



Abstract We consider the nonparametric estimation of spectral densities for second-order stationary random fields on a d -dimensional lattice. We discuss some drawbacks of standard methods and propose modified estimator classes with improved bias convergence rate, emphasizing the use of kernel methods and the choice of an optimal smoothing number. We prove the uniform consistency and study the uniform asymptotic distribution when the optimal smoothing number is estimated from the sampled data.

Keywords Spatial data · Spectral density · Smoothing number · Uniform asymptotic distribution · Bootstrap

Mathematics Subject Classification (2000) Primary 62M30 · Secondary 62M15 · 62G20

1 Introduction

The estimation of the power spectrum for random fields on a d -dimensional lattice is relevant for many purposes, including specification and testing of parametric models, detecting anisotropies and hidden periodicities, signal extraction from noisy random fields, interpolation, prediction, and smoothing. It is also useful to obtain a more sparse decomposition of a digital image, requiring less storage space. Spatial spectral

I wish to thank Professor C. Velasco and two anonymous referees for their helpful comments and suggestions and Professor P.M. Robinson for introducing me with the topic. This research has been supported by a Marie Curie Grant, Mobility 11, of the European Commission, reference number FP6-2004-505469.

J.M. Vidal-Sanz (✉)

Department of Business Economics, Universidad Carlos III de Madrid, C/ Madrid 126,
28903 Getafe (Madrid), Spain
e-mail: jvidal@emp.uc3m.es

methods have been applied to ecological data (e.g., Reshaw 1984, and Reshaw and Ford 1983), earth sciences (Agterberg 1967), astronomy (Abramenko et al. 2001), and meteorology (Barry and Perry 1973), among others.

This paper is concerned with nonparametric estimation of the spectral density for spatial processes. We discuss some drawbacks in the current estimation methods. The bias of nonparametric estimators based on Whittle's (1954) periodogram achieve a low convergence rate due to the "edge effects," whilst the smoothed periodogram based on Guyon's (1982) periodogram can present consistency problems. We overcome these problems by smoothing the modified periodogram introduced by Robinson and Vidal-Sanz (2006). We focus on kernel estimators, for which we consider the choice of an optimal smoothing number. Furthermore, the uniform consistency and uniform asymptotic distribution are established when the optimal smoothing number is estimated from the data (see Theorem 3). The uniform asymptotic distribution result has also interest in time series context, complementing Robinson's (1991) uniform consistency result for automatic estimation. Finally, we present a consistent Bootstrap method for automatic estimation of the smoothing number.

Consider a real second-order stationary stochastic process $\{X_t : t \in \mathbb{Z}^d\}$ on a d -dimensional lattice, where $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, with first moments $E[X_t] = \mu$ and $\gamma_l = \text{Cov}[X_t, X_{t+l}]$. We assume that there exists an integrable spectral density $f(\lambda) \geq 0$ on $\Pi^d = [-\pi, \pi]^d$ such that $\gamma_l = \int_{\Pi^d} e^{il \cdot \lambda} f(\lambda) d\lambda$ with $l \cdot \lambda = \sum_{r=1}^d \lambda_r l_r$ and $f(\lambda) = (2\pi)^{-d} \sum_{l \in \mathbb{Z}^d} \gamma_l e^{-il \cdot \lambda}$ (this is why f is also called the power spectrum). The spectral density can be periodically extended to \mathbb{R}^d . A sufficient condition for the existence of f is that $\sum_{l \in \mathbb{Z}^d} |\gamma_l| < \infty$; this also implies that $f \in C(\Pi^d)$ and it obeys the Lipschitz condition $f \in \text{Lip}(\alpha)$ for any $\alpha < 1/2$, where $f \in \text{Lip}(\alpha)$ means that $\sup_{0 < \|h\| \leq \delta} \|f(\lambda) - f(\lambda+h)\|_{\infty} = o(\delta^\alpha)$ as $\delta \downarrow 0$ with $\|f\|_{\infty} = \sup_{\lambda \in \Pi^d} |f(\lambda)|$. Under the stronger condition $\sum_{l \in \mathbb{Z}^d} (1 + \|l\|_1^q) |\gamma_l| < \infty$ for some integer $q \geq 1$, where $\|l\|_1 = \sum_{r=1}^d |l_r|$, we have that $f \in C^q(\Pi^d)$.

In spatial data, it is customary to take the beginning data situated at the origin (or at one), but sometimes data are centered elsewhere, and asymptotic could require that the sample increases in all directions of space. Therefore, without loss of generality, we consider the estimation of the spectral density when X_t is observed on a rectangular set

$$N = \{t \in \mathbb{Z}^d : -n_r^L \leq t_r \leq n_r^U, r = 1, \dots, d\},$$

where $n_r^L, n_r^U \in \mathbb{Z}$ and $-n_r^L \leq n_r^U$ for $r = 1, \dots, d$. Then define $n_r = n_r^L + n_r^U + 1$ and $n = \prod_{r=1}^d n_r$ the cardinal of N . Following Robinson and Vidal-Sanz (2006), for the asymptotic regime, we regard $n_r = n_r(n)$ as a function of the total number of observations n , which is the basic index for asymptotic results; and we require that n_r increases for all r , introducing the following assumption:

A.1 For all sufficiently large n , there exist $\xi > 0$ and $c > 0$ such that

$$n_r(n) > c_1 n^\xi.$$

Since $\sum_{r=1}^d n_r(n)^{-1} \leq d(\prod_{r=1}^d n_r^{-1})^{1/d} = dn^{-1/d}$, we have that $\xi \leq 1/d$, where the equality is attained when all $n_r(n)$ increase at the same rate $n^{1/d}$. This specification covers many possibilities. For example, we could set $n_r^L = -1$ for $r = 1, \dots, d$

and therefore consider the standard unilateral sample case $N = \times_{r=1}^d \{1, \dots, n_r^U\}$ with $n = \prod_{r=1}^d n_r^U$ and $n_r^U \rightarrow \infty$. The spatial statistics literature focuses on this case, but spatial samples could generally increase in one or several directions. For example, we can observe a symmetric sample with $n_r^L = -n_r^U$, $n_r^U \geq 0$, for all coordinates, so that $N = \times_{r=1}^d \{-n_r^U, \dots, n_r^U\}$ with $n = \prod_{r=1}^d (2n_r^U + 1)$.

For any $l \in \mathbb{Z}^d$, let us define $N(l) := \{t \in N : t + l \in N\}$ with cardinal $n(l) := \prod_{r=1}^d (n_r - |l_r|)$, the unbiased covariance estimator

$$c_{n,l}^* = \frac{1}{n(l)} \sum_{t \in N(l)} (X_t - \hat{\mu}_n)(X_{t+l} - \hat{\mu}_n)$$

with $\hat{\mu}_n = n^{-1} \sum_{t \in N} X_t$, the biased covariance estimator $c_{n,l} = w(l) c_{n,l}^*$ with $w(l) = n(l)n^{-1} = \prod_{r=1}^d (1 - |l_r|/n_r)$, and the discrete Fourier transform

$$d_n(\lambda) = (n(2\pi)^d)^{-1/2} \sum_{t \in N} (X_t - \hat{\mu}_n) e^{-it \cdot \lambda}.$$

Whittle (1954) introduced the spatial periodogram in the context of unilateral samples. The spatial periodogram

$$I(\lambda) = |d_n(\lambda)|^2 = (2\pi)^{-d} \sum_l \prime c_{n,l} e^{-il \cdot \lambda} = (2\pi)^{-d} \sum_l \prime w(l) c_{n,l}^* e^{-il \cdot \lambda},$$

where $\sum_l \prime$ denotes the sum for $l \in \mathbb{Z}^d$ such that $|l_r| \leq n_r - 1$, $r = 1, \dots, d$, is asymptotically unbiased for $f(\lambda)$. But the variance of $I(\lambda)$ does not tend to zero, as it can be anticipated, and some smoothing is required. Henceforth, we will use the discrete frequencies $\omega_{j,n} = (2\pi j_1/n_1, \dots, 2\pi j_d/n_d)$ for all $j \in J_n$, where the set $J_n = \times_{r=1}^d \{0, \dots, n_r - 1\}$ has cardinal n . The numerical effort required to compute $I(\omega_{j,n})$ can be reduced by using the planar Fast Fourier Transform, see Reshaw and Ford (1983) for a discussion.

The spatial literature has discussed the nonparametric spectral density estimation for random fields with samples spreading in one direction ($n_r^L \equiv -1$, $n_r^U \rightarrow \infty$), see, e.g., Priestley (1964), Rozanov (1967), Brillinger (1970), Rosenblatt (1985), Ivanov and Leonenko (1986), Žurbenko (1986), Heyde and Gay (1993), and Leonenko (1999), among others. The basic theory is a straightforward generalization from time series. One of the most simple estimators is the class of *smoothed periodogram* estimators

$$\hat{f}(\lambda) = \frac{1}{n} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda) I(\omega_{j,n}),$$

where M is a smoothing number. The weight functions $\{K_M\} \subset L_1(\mathbb{R}^d)$ are symmetric, continuous, and periodical with periodicity $[0, 2\pi]^d$, and, as the smoothing parameter $M \rightarrow \infty$, the functions $K_M \rightarrow \delta_0$ (where δ_0 is the periodic extension of the Dirac delta generalized function at 0). The consistency requires M depending on n with $M_n \rightarrow \infty$ at a rate sufficiently slow to ensure that the variance of \hat{f} tends

to zero. Another popular class of spectral density estimators, known as *lag windowed* estimators, is defined by

$$\tilde{f}(\lambda) = (2\pi)^{-d} \sum_l \prime k_M(l) w(l) c_{n,l}^* e^{-il \cdot \lambda}, \quad (1)$$

where $k_M(l)$ is the lag window satisfying $k_M(l) = k_M(-l) \leq k_M(0) = 1$, and the parameter M plays the role of a smoothing number. It is possible to consider different kinds of smoothing numbers. When $M \in \mathbb{N}^d$ and $k_M(l) = 0$ for $|l_r| \geq M_r$ and $r = 1, \dots, d$, the parameters M are called lag numbers. The smoothing numbers could be positive-definite matrices $M \in \mathbb{R}^{d \times d}$ such that $k_M(l) = k(M^{-1}l)$ with $|k(l)| \leq k(0) = 1$ for all l and $k(l) = 0$ for $|l| \geq 1$. For diagonal matrices, the vector $\text{diag}(M)$ can be regarded as lag numbers. Lag windowed and smoothed periodogram estimators can be related. Introducing $K_M(u) = (2\pi)^{-d} \sum_l \prime k_M(l) e^{-il \cdot u}$, we can express lag windowed estimators as

$$\tilde{f}(\lambda) = \int_{\Pi^d} \left((2\pi)^{-d} \sum_l \prime k_M(l) e^{-il \cdot (\lambda - u)} \right) I(u) du = \int_{\Pi^d} K_M(u - \lambda) I(u) du,$$

where $k_M(l) = \int_{\Pi^d} e^{il \cdot \lambda} K_M(\lambda) d\lambda$. Thus, \hat{f} can be thought of as a numerical integration approximation to \tilde{f} .

For any of these estimator classes, the consistency can be established much as in the time series literature. Unfortunately, the spatial density estimators previously discussed are exposed to a low bias convergence rate, inherent in the Whittle spatial periodogram. As $E[I(\lambda)]$ is the Cesaro sum of the multiple Fourier series of f (see, e.g., Zygmund 1959, Vol. II, Chap. XVII),

$$E[I(\lambda)] = (2\pi)^{-d} \sum_l \prime w(l) \gamma_l e^{-il \cdot \lambda} = \int_{\Pi^d} F_n(u - \lambda) f(u) du,$$

where $F_n(u) = \prod_{r=1}^d (2\pi n_r)^{-1} (\sin\{n_r u_r / 2\} / \sin\{u_r / 2\})^2$ is the multivariate Fejer kernel. Let us consider $\omega(f, \delta) = \sup_{0 < \|h\| \leq \delta} \|f(\lambda) - f(\lambda + h)\|_\infty$. As a consequence of the Korovkin theorem (see, e.g., Korovkin 1960), we have that, as $n \rightarrow \infty$,

$$\|E[I(\lambda)] - f\|_\infty \leq 2\omega(f, \delta_n) = o(\delta_n^{1/2})$$

for all $f \in \text{Lip}(\alpha)$ with $\alpha > 1/2$, where

$$\begin{aligned} \delta_n &= \int_{\Pi^d} F_n(u - \lambda) \|u\| du \leq K \prod_{i=1}^d \int_{-\pi}^{\pi} \frac{1}{2\pi n_i} \left(\frac{\sin\{n_i u_i / 2\}}{\sin\{u_i / 2\}} \right)^2 \left(\sum_{r=1}^d |u_r| \right) du \\ &= K \sum_{r=1}^d \int_{-\pi}^{\pi} \frac{1}{2\pi n_r} \left(\frac{\sin\{n_r u_r / 2\}}{\sin\{u_r / 2\}} \right)^2 |u_r| du = O\left(\sum_{i=1}^d n_r^{-1} \right), \end{aligned}$$

which by Assumption A.1 is of order not less than $n^{-1/d}$, and the uniform bias rate of I can be lower than $o(1/\sqrt{n})$ for $d > 1$. The basic reason for the low convergence

rate is the edge effect noticed by Guyon (1982). For a fixed l , as all $n_r \rightarrow \infty$, the bias $|E[c_{n,l}] - \gamma_l|$ is of order $\sum_{r=1}^d n_r^{-1} \geq dn^{-1/d}$. Thus, for a continuous integrable kernel K ,

$$\left| \frac{1}{n} \sum_{j \in J_n} K(\omega_{j,n} - \lambda) (E[I(\omega_{j,n})] - f(\omega_{j,n})) \right| = o(n^{-\xi/2})$$

by Assumption A.1. Therefore, the uniform rate of convergence is $o(1/\sqrt{n})$ only for $d = 1$ but can be significantly slower for $d > 1$.

To avoid the edge effect, Guyon (1982) introduced the modified periodogram with unbiased covariances

$$I_*(\lambda) = (2\pi)^{-d} \sum_l c_l^* e^{-il\lambda}$$

for unilateral samples. Note that $I_*(\lambda)$ is not necessarily a nonnegative function, and $E[I_*]$ is the multiple Fourier series of f . Although there are infinitely many continuous functions f the Fourier series of which diverge to infinity (see, e.g., Rudin 1974, and Vidal-Sanz 2005), $\|E[I_*(\lambda)] - f(\lambda)\|_\infty \rightarrow 0$ if f is a continuous function with bounded variation on Π^d . The modified periodogram I_* can be smoothed to estimate the spectral density f when it is enough regular. Politis and Romano (1996) suggested to use unbiased autocovariances in spectral density estimation. The *lag windowed* estimator based on I_* is

$$\tilde{f}_*(\lambda) = \int_{\Pi^d} K_M(u - \lambda) I_*(u) du = (2\pi)^{-d} \sum_l k_M(l) c_{n,l}^* e^{-il\lambda},$$

similar to (1) with lag window $\{k_M(l)/w(l)\}$.

The theoretical properties of I_* have been criticized by Robinson and Vidal-Sanz (2006) in the context of Whittle estimation, due to the presence of aliasing problems. This problem can also be found in *smoothed periodogram* estimators; it suffices to consider the weight function $K_M(\lambda) = (2\pi)^{-d} \sum_l k_M(l) e^{-il\lambda}$. Applying Hannan's (1973) argument, we have that

$$\hat{f}_*(\lambda) = \frac{1}{n} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda) I_*(\omega_{j,n}) = \sum_l k_M(l) (c_l^* + c_{l \pm n}^*), \quad (2)$$

where $c_{l \pm n}^* = 0$ for $l = 0$ and $c_{l \pm n}^* = c_{n-l}^*$ for $l \neq 0$. The right-hand side of (2) is equal to

$$= (2\pi)^{-d} \int_{\Pi^d} K_M(\omega - \lambda) I_*(\omega) d\omega + \sum_l k_M(l) c_{l \pm n}^*,$$

where $c_{l \pm n}^*$ is composed of at most $n - l$ terms of the form $X_l X_{n-l+i}$ divided by l , which does not converge to zero (e.g., $c_{n-1}^* = X_1 X_n$). Although $k_M(l) \rightarrow 0$, if this convergence is not uniform in l , a smoothed periodogram based on I_* could be inconsistent or, in the best case, the rate of convergence could be too slow. By contrast, in the Whittle periodogram, $c_{l \pm n} = O_p(n^{-1})$, and the ‘‘aliasing’’ of lags does not generate the inconsistency, as proved by Hannan (1973).

Dalhaus and Künsch (1987) proposed to use a periodogram I_T the covariances of which use tapered data, using this periodogram for Whittle estimation of parametric models. They show that, for $d \leq 3$, if the taper uses an appropriate bandwidth, the estimated parameters are consistent with rate \sqrt{n} . Robinson (2007) suggested to use tapered periodograms in spectral density estimation. But for lag windowed spectral estimators based on a such periodogram, it would be required to choose a taper, a bandwidth, and a smoothing number, thus introducing too much ambiguity in the estimation.

In this paper, we present a modified spectral density estimator which is affected neither by the aliasing nor by the edge effect. In Sect. 2, we introduce the modified estimators, focusing on kernel estimators. We also consider the optimal smoothing number for the integrated mean-square loss function, which is infeasible and has to be estimated from the sample data. We also discuss the issue of spatial sampling intervals. Section 3 contains the main theoretical results. For a stochastic smoothing number, we prove the uniform consistency and pointwise asymptotic normality of modified kernel estimators. In Sect. 4, we consider the consistency of plug-in and Bootstrap estimators of the optimal smoothing number. The proofs can be found in Vidal-Sanz (2007).

2 Modified spectral density estimators

To avoid the aliasing problems in I_* , Robinson and Vidal-Sanz (2006) introduced the truncated periodogram

$$I_g(\lambda) = (2\pi)^{-d} \sum_{\substack{l \in \mathbb{Z}^d: |l_r| \leq g(n_r) \\ r=1, \dots, d}} c_l^* e^{-il \cdot \lambda},$$

where g is a function satisfying:

A.2 g is a positive, integer-valued, and monotonically increasing function such that $g(x) \rightarrow \infty$ as $x \rightarrow \infty$ and, for some $c_2 \in (0, 1)$, $g(x) \leq c_2 x$ for all $x > 0$.

For example, we can take $g(x) = [\alpha x]$ with $\alpha \in (0, 1)$ and $[\cdot]$ the integer part; in practice, this means that we consider a trimmed summation of elements l with coordinates $|l_r| \leq \alpha n_r$. The advantage of this approach is that the parameters in the function g do not play an asymptotic effect, by contrast with tapering methods. Some finite sample experiments can be found in Robinson and Vidal-Sanz (2006).

Robinson and Vidal-Sanz (2006) prove that $\|E[I_g] - f\| = o(n^{-1/2})$ under appropriate assumptions on the covariance function (A.3 and A.4 in Sect. 3); and when averaged over discrete frequencies, the modified periodogram I_g is immune to aliasing problems affecting I_* . Here we propose the class of *modified smoothed periodogram* estimators

$$\hat{f}^g(\lambda) = \frac{1}{n} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda) I_g(\omega_{j,n}) \quad (3)$$

and the class of *modified lag windowed* estimators

$$\tilde{f}^g(\lambda) = \int_{\Pi^d} K_M(u - \lambda) I_g(u) du = (2\pi)^{-d} \sum_{|l_r| \leq g(n_r)} \cdots \sum_{r=1, \dots, d} k_M(l) c_l^* e^{-il \cdot \lambda} \quad (4)$$

with $K_M(u) = (2\pi)^{-d} \sum_{r=1}^d \sum_{|l_r| \leq g(n_r)} k_M(l) e^{-il \cdot u}$ and $k_M(l) = \int_{\Pi^d} e^{il \cdot \lambda} \times K_M(\lambda) d\lambda$. Both estimators \hat{f}^g and \tilde{f}^g are similar to the classical ones, but using I_g (instead of I or I_*) it is possible to establish the uniform consistency and derive the uniform weak distribution under appropriate conditions. As I_* , I_g can also take negative values for some frequencies, and so do \hat{f}^g and \tilde{f}^g . Although negative frequency estimations are unlikely for large samples, we can vanish the estimator for frequencies with negative estimated power spectra by taking $\hat{f}^{g+}(\lambda) = \max\{0, \hat{f}^g(\lambda)\}$, i.e., the L_1 -projection of \hat{f}^g onto the positive cone. \tilde{f}^{g+} is defined analogously.

A rigorous treatment of the asymptotic theory is given in Sect. 3, but some heuristic arguments are presented in this section. Proceeding much as in the time-series literature, under appropriate conditions, the estimator \tilde{f}^g roughly satisfies

$$E[\tilde{f}^g(\lambda)] = \int_{\Pi^d} K_M(u - \lambda) f(u) du + o(n^{-1/2}),$$

$$\text{Cov}[\tilde{f}^g(\lambda), \tilde{f}^g(\theta)] \approx \frac{(2\pi)^d}{n} \int_{\Pi^d} K_M(u - \lambda) K_M(u - \theta) f(u)^2 du,$$

and \hat{f}^g exhibits a similar behavior, as the aliasing of lags does not affect the modified smoothed periodogram.

Applying the Korovkin theorem, one can prove that \tilde{f}^g is asymptotically unbiased for integrable and continuous f , and a Lipschitz assumption can be used to obtain a convergence rate. If $f \in C^r(\Pi^d)$, taking the Taylor expansion of $f(u + \lambda)$ in λ , we obtain that

$$\begin{aligned} E[\tilde{f}^g(\lambda)] - f(\lambda) &= \int_{\Pi^d} K_M(u) (f(u + \lambda) - f(\lambda)) du \\ &= \sum_{j=1}^{r-1} \sum_{\|v\|_1=j} \frac{D^v f(\lambda)}{v!} \int_{\Pi^d} K_M(u) u^v du \\ &\quad + r \sum_{\|v\|_1=r} \frac{1}{v!} \int_{\Pi^d} \int_0^1 (1-t)^{r-1} D^v f(\lambda + tu) K_M(u) u^v dt du. \end{aligned}$$

We say that the family $\{K_M(u)\}$ is of order r if, for all M , we have $\int_{\Pi^d} K_M(u) u^v du = 0$ for $1 \leq \|v\|_1 < r$ and $\int_{\Pi^d} \|u\|^r |K_M(u)| du < \infty$; this implies that the bias convergence rate to zero equals the rate of the remaining term, namely, $O(\int_{\Pi^d} \|u\|^r |K_M(u)| du)$, uniformly in frequency. In particular, the symmetry $k_M(l) = k_M(-l)$ implies that $\int_{\Pi^d} u K_M(u) du = 0$, and the bias rate is $O(\int_{\Pi^d} \|u\|^2 |K_M(u)| du)$ for $f \in C^2(\Pi^d)$. In some particular cases (e.g., kernel estimators), it is easy to obtain orders higher than 2, but it is not for general estimators.

Delgado and Vidal-Sanz (2001) present a general methodology for obtaining families $\{K_M(\cdot)\}$ with higher orders.

Regarding the covariance structure, if K_M is supported on a closed neighborhood around the origin, the covariance tends to zero for $\lambda \neq \theta$, and the variance satisfies

$$\begin{aligned} \text{Var}[\tilde{f}^g(\lambda)] &\approx \frac{(2\pi)^d}{n} \int_{\Pi^d} K_M(u)^2 f(u + \lambda)^2 du \\ &\approx f(\lambda)^2 \frac{(2\pi)^d}{n} \int_{\Pi^d} K_M(u)^2 du. \end{aligned}$$

This approximation is accurate for M , n large, or f flat around λ . The estimator will be mean-square consistent when $\int_{\Pi^d} K_{M_n}(u)^2 du = o(n)$.

Several functional norms $\|\cdot\|$ can be used to study the global convergence $\|\hat{f}^g - f\| \rightarrow 0$ in probability, i.e., different function spaces can be considered. Perhaps the most popular choices are $C(\Pi^d)$ endowed with the supremum norm $\|f\|_\infty = \sup_{\lambda \in \Pi^d} |f(\lambda)|$ and the space $L_2(\mu)$ for some Borel measure μ on Π^d endowed with the mean-square norm $\|f\|_{L_2(\mu)} = (\int_{\Pi^d} |f(\lambda)|^2 \mu(d\lambda))^{1/2}$, where the Lebesgue measure is frequently taken. Both are complete and separable Banach spaces, and $C(\Pi^d)$ is dense in $L_2(\mu)$. The uniform consistency is stronger than the L_2 consistency on Π^d and it will be considered in Sect. 3.

2.1 Kernel estimators

Perhaps the most relevant methods are (*modified*) *kernel estimators*, and the rest of the paper is focused on them. There are two alternative approaches to introduce kernel estimators. In the first one, kernel estimators are a class of smoothed periodograms (3), whilst, in the second one, they are lag windowed methods (4). The distinctive aspect of kernel methods is that the kernel K_M is defined by

$$K_M(u) = \det(M) \sum_{l \in \mathbb{Z}^d} K(M(u + 2\pi l)), \quad u \in \Pi^d,$$

for a kernel function $K \in L_1(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} K(u) du = 1$. The kernel K can be defined as the product of univariate kernels $K(u) = \prod_{r=1}^d K_r(u_r)$ with $\{K_r\} \subset L_1(\mathbb{R})$. The smoothing number M_n is a sequence of symmetric positive definite matrices with $M_n \rightarrow \infty$ and $\det(M_n)/n \rightarrow 0$. The kernel lag window $k_M(l)$ verifies

$$\begin{aligned} k_M(l) &= \int_{\Pi^d} e^{il \cdot u} K_M(u) du = \det(M) \sum_{l \in \mathbb{Z}^d} \int_{\Pi^d} e^{il \cdot u} K(M(u + 2\pi l)) du \\ &= \det(M) \int_{\mathbb{R}^d} e^{il \cdot u} K(Mu) du = \int_{\mathbb{R}^d} e^{il \cdot M^{-1}u} K(u) du = k(M^{-1}l) \end{aligned}$$

with $k(x) = \int_{\mathbb{R}^d} e^{ix \cdot u} K(u) du$. Therefore, if $k \in L_2(\mathbb{R}^d)$, applying Parseval's equality, we have

$$\int_{\Pi^d} K_M(u)^2 du = \frac{\det(M)}{(2\pi)^d} \int_{\mathbb{R}^d} k(u)^2 du$$

and $\int_{\Pi^d} K_{M_n}(u)^2 du = o(n)$ as $\det(M_n)/n \rightarrow 0$.

Let us consider the matrix norm $\|M\| = (\text{megv}(M'M))^{1/2}$, where megv means the maximum eigenvalue. We say that K is a kernel of order q if $\int_{\Pi^d} K(u)u^v du = 0$ for $1 \leq \|v\|_1 < q$ and $\int_{\Pi^d} |K(u)|\|u\|^q du < \infty$. The q -order property ensures that, for $f \in C^q(\Pi^d)$, the spectral density bias is $O(\|M_n\|^{-q})$ uniformly in frequency. This high-order rate is a relevant property in order to ensure that the bias tends to zero at rate $o(n^{-1/2})$. Note that K is of order $q = 2$ whenever it is even and $\int_{\Pi^d} |K(u)|\|u\|^2 du < \infty$. For K to be of order $q > 2$, it is necessary that K takes negative values. The q -order kernel property can be stated by the requirement that $k(x)$ is q -times continuously differentiable at zero with $D^v k(x)|_{x=0} = 0$ for all integer vectors $0 < \|v\|_1 < q$. Since $k(x) \leq k(0) = 1$, taking into account the Taylor expansion definition, the last condition can be equivalently expressed by the following condition:

$$\lim_{x \rightarrow 0} \frac{1 - k(x)}{\|x\|^q} = k_q$$

for some finite constant k_q . The extreme case $q = \infty$ is often identified with the ‘‘flat-top’’ kernels considered by Politis and Romano (1996).

The variance of kernel estimators is $O(\det(M_n)/n)$, and the square bias is at best $O(\|M_n\|^{-2q})$; both rates are satisfied uniformly in frequency. If $M_n = m_n S$ with m_n scalar and S a symmetric positive definite matrix, then the mean-square error of kernel estimation is $O(m_n^{-2q} + m_n^d/n)$ uniformly in frequency, and the rate of convergence is made as fast as possible by taking $m_n = O(n^{1/(2q+d)})$, with associated mean square error $O(n^{-\frac{2q}{2q+d}})$. In particular, when $q = d = 2$, the optimal rate is $m_n = O(n^{1/6})$, which suggests that we could take $M_n = S n^{1/6}$ for a matrix S , and the associated mean-square error is $O(n^{-4/6})$. The curse of dimensionality can be observed, as the mean-square error rate $n^{2q/(2q+d)}$ decreases exponentially when the dimension d increases, implying that, for high dimensions, the sample size n required for accurate estimations should be increasingly large. In space-temporal context, we rarely find dimensions $d > 4$, and actual sample sets are usually large enough to avoid concerns about this issue.

2.2 Kernel optimal smoothing numbers

The choice of the parameter S is crucial to deal with the trade-off effects between the bias and variance in finite samples, and it should be based on some objective loss function. Different loss functions lead to different optimal parameters S^* that usually depend on the unknown f , but there is no a universally optimal parameter for all loss criteria. A relevant and manageable loss function is the integrated mean-square error with respect to the weight measure μ

$$\begin{aligned} \text{IMSE}(M, n, \mu) &= E[\|\tilde{f}^g - f\|_{L_2(\mu)}^2] = \int_{\Pi^d} E[|\tilde{f}^g(\lambda) - f(\lambda)|^2] \mu(d\lambda) \\ &= \int_{\Pi^d} E[|\tilde{f}^g(\lambda) - E[\tilde{f}^g(\lambda)]|^2] \mu(d\lambda) \\ &\quad + \int_{\Pi^d} |E[\tilde{f}^g(\lambda)] - f(\lambda)|^2 \mu(d\lambda) \end{aligned}$$

by Fubini's theorem. If $f \in C^2(\Pi^d)$ and K is of order 2, the bias is

$$\begin{aligned} \text{Bi}[\tilde{f}^g(\lambda)] &= \int_{\Pi^d} K(u)(f(\lambda + M_n u) - f(\lambda)) \\ &= \int_{\Pi^d} K(u) \left(u' M_n \nabla f(\lambda) du + \frac{1}{2} u' M_n^{-1'} \frac{\partial^2}{\partial \lambda \partial \lambda'} f(\lambda) M_n^{-1} u \right) du \\ &\quad + o(\|M_n\|^{-2}) \\ &= \frac{1}{2} \text{Tr} \left\{ \int_{\Pi^d} u u' K(u) du \cdot M_n^{-1'} \frac{\partial^2}{\partial \lambda \partial \lambda'} f(\lambda) M_n^{-1} \right\}^2 \mu(d\lambda) \\ &\quad + o(\|M_n\|^{-2}), \end{aligned}$$

where the $o(\|M_n\|^{-2})$ term is uniform in λ , and Tr denotes the trace of a square matrix.

Proceeding heuristically (a precise treatment is presented in Sect. 3), we have that

$$\text{Var}[\tilde{f}^g(\lambda)] \approx \frac{\det(M_n)}{n} \kappa^2 f(\lambda)^2$$

with $\kappa^2 = \int_{\mathbb{R}^d} k(u)^2 du$, and the $o(\cdot)$ term is uniform in λ ; taking $M_n = m_n S$, we have

$$\text{IMSE}(m_n S, n, \mu) \approx \frac{m_n^d \det(S)}{n} \kappa^2 F^2 + \frac{C_K^2}{4m_n^4} \int_{\Pi^d} \left| \text{Tr} \left\{ (SS')^{-1} \frac{\partial^2}{\partial \lambda \partial \lambda'} f(\lambda) \right\} \right|^2 \mu(d\lambda),$$

where $C_K = \int_{\Pi^d} u u' K(u) du$ and $F^2 = \int_{\Pi^d} f(\lambda)^2 \mu(d\lambda)$. If we use the optimal rate for $q = 2$, i.e., $m_n = n^{1/(4+d)}$, then, for n large,

$$\begin{aligned} n^{4/(4+d)} \text{IMSE}(S n^{1/(4+d)}, n, \mu) \\ \approx \det(S) \kappa^2 F^2 + \frac{C_K^2}{4} \int_{\Pi^d} \left| \text{Tr} \left\{ (SS')^{-1} \frac{\partial^2}{\partial \lambda \partial \lambda'} f(\lambda) \right\} \right|^2 \mu(d\lambda). \end{aligned}$$

The right-hand side can be minimized in S by taking

$$S_0^* = \left(\frac{4C_K^2}{\kappa^2 F^2} \int_{\Pi^d} \left(\frac{\partial^2}{\partial \lambda \partial \lambda'} f(\lambda) \right)^2 \mu(d\lambda) \right)^{1/4+d}.$$

Therefore, we do not use the same bandwidth in each dimension of the frequency space, but rather a general elliptically shaped kernel at a particular rotation controlled by $(S_0^* S_0^{*'})$. Similar arguments can be applied for $E[\|\hat{f}^g - f\|_{L_2(\mu)}^2]$. Higher-order kernels can be considered, but f should satisfy higher differentiability requirements. In all the cases, the optimal value is a function $S_0^* = S^*(f)$ of the unknown f .

Though $S_0^* = S^*(f)$ is infeasible, usually it can be estimated from the data by a plug-in procedure, some cross-validation method, or Bootstrap. The plug-in procedure takes a consistent pilot estimation $\tilde{f}_{M_0}^g$ and estimates $\hat{S}^* = S^*(\tilde{f}_{M_0}^g)$. For example, when $f \in C^3(\Pi^d)$ and some regularity conditions are satisfied, we can use a

kernel pilot, since $\partial^2 \tilde{f}^g / \partial \lambda \partial \lambda'$ is consistent with respect to $\partial^2 f / \partial \lambda \partial \lambda'$. The plug-in procedure can be iterated. Cross-validation methods are popular in time-series analysis, see Beltrao and Bloomfield (1987) and Robinson (1991, Sect. 5) and they can be extended to deal with spatial data. However, in this paper, we will focus on Bootstrap methods. Our approach is different from that of Frank and Härdle (1992) time-series bootstrap method for kernel spectral estimators based on a Studentized periodogram. See Sect. 4 for details.

Summarizing, nonparametric estimation of power spectrum requires the choice of an appropriate smoothing number M_n . The choice of an optimal smoothing number entails the choice of a loss function leading to some optimal smoothing number, usually infeasible though it can be estimated from the sampled data. As a consequence, the smoothing number M_n should be allowed to depend on the data, provided that $\det(M_n)/n \rightarrow_p 0$ and $M_n \rightarrow_p \infty$, as required for the mean-square consistency.

2.3 Sampling effects

Earth sciences often collect data from a continuous phenomena at regular intervals, using fixed monitoring points. Consider a real second-order stationary stochastic process $\{X_t : t \in \mathbb{R}^d\}$ with continuous spatial index with spectral density $f \in C(\mathbb{R}^d)$ and covariances $\gamma_l = \int_{\mathbb{R}^d} f(\lambda) e^{il \cdot \lambda} d\lambda$. Assume that the sampling interval for each coordinate is $\Delta = (\Delta_1, \dots, \Delta_d)^T$. For any $t \in \mathbb{Z}^d$, define $t \otimes \Delta = (t_1 \Delta_1, \dots, t_d \Delta_d)^T$ and $\Pi_\Delta^d = \prod_{r=1}^d [-\pi/\Delta_r, \pi/\Delta_r]$. The upper limit of the interval, $(\pi \Delta_1^{-1}, \dots, \pi \Delta_d^{-1})$, is known as the Nyquist or folding frequency. Then the sampled process $\{X_{t \otimes \Delta} : t \in \mathbb{Z}^d\}$ has the spectral density f_Δ given by the folding formula

$$f_\Delta(\lambda) = \sum_{j \in \mathbb{Z}^d} f(\lambda + \omega_{j, \Delta}),$$

where $\omega_{j, \Delta} = (2\pi j_1/\Delta_1, \dots, 2\pi j_d/\Delta_d)$ are called alias, and $\lambda \in \Pi_\Delta^d$. A peak on the spectrum f_Δ observed at frequency λ can be caused by an aliased frequency $\omega_{j, \Delta}$, unless f possesses no components with frequency greater than the Nyquist frequency, i.e., $f_\Delta(\lambda) = f(\lambda)$ for $\lambda \in \Pi_\Delta^d$. Note that $\gamma_{l \otimes \Delta} = \int_{\Pi_\Delta^d} f_\Delta(\lambda) e^{i(l \otimes \Delta) \cdot \lambda} d\lambda$ for all $l \in \mathbb{Z}^d$ and $f_\Delta(\lambda) = \prod_{r=1}^d (\Delta_r/2\pi) \sum_{l \in \mathbb{Z}^d} \gamma_{l \otimes \Delta} e^{-i(l \otimes \Delta) \cdot \lambda}$. Using the observed data $\{X_{t \otimes \Delta} : t \in N\}$, a modified nonparametric estimator of f_Δ can be defined similarly to the case of unit sampling distance, i.e., smoothing the modified periodogram

$$I_{\Delta g}(\lambda) = \prod_{r=1}^d (\Delta_r/2\pi) \sum_{\substack{l \in \mathbb{Z}^d: |l_r| \leq g(n_r) \\ r=1, \dots, d}} c_{l \otimes \Delta}^* e^{-i(l \otimes \Delta) \cdot \lambda}.$$

The presented approach is valid to study the statistical behavior of the process on the regular sampling net, but something can be inferred about the continuous process when data are densely collected. Since $f(\lambda) \rightarrow 0$ as $\|\lambda\| \rightarrow \infty$ for an integrable f , for a sufficiently small Δ , there are no appreciable components in f with frequencies higher than the Nyquist frequency, and the estimator \hat{f}_Δ can be used to infer approximately the behavior of f . The error decreases slowly only when f has heavy tails, i.e., when γ_l presents nonsmooth features.

3 Main results

This section is devoted to the uniform consistency and uniform asymptotic distribution of modified kernel spectral density estimators with multilateral samples. To derive the asymptotic theory, we will assume a linear representation, introducing the following assumption:

A.3 *The spatial process $\{X_t\}_{t \in \mathbb{Z}^d}$ follows a second-order stationary random field with linear representation*

$$X_t = \mu + \sum_{j \in \mathbb{Z}^d} \beta_j \varepsilon_{t-j},$$

where $\sum_{j \in \mathbb{Z}^d} |\beta_j| < \infty$, and $\{\varepsilon_j\}$ are independent identically distributed random variables with zero mean, variance σ_ε^2 , and fourth-order cumulant $\kappa_\varepsilon < \infty$.

Other approaches have been pursued in the literature. For example, we can assume conditions on the existence, stationarity, and summability of higher-order cumulants of $\{X_t\}$, using arguments related to Brillinger (1981). But for the estimation of second order spectra, it is not really necessary to involve conditions on higher moments. Markovian assumptions or m -dependence conditions can be also considered to derive asymptotic results, but spatial correlations often decay slowly (see, e.g., Ripley 1988, p. 3). Mixing conditions are often used, see Doukhan (1994) for a review. Perhaps, Bolthausen's (1982) central limit theorem for α -mixing random fields is the most popular method. Linear processes, as described in A.3, are often used to justify the α -mixing assumption for X_t , under the requirement that the probability density function of ε_t satisfies a Lipschitz condition. We avoid this requirement by following the martingale difference approach based on A.3. We also assume that:

A.4 *The spatial process $\{X_t\}_{t \in \mathbb{Z}^d}$ follows a second-order stationary random field with autocovariance function $\gamma_l = \text{Cov}[X_0, X_l]$ satisfying*

$$\sum_{l \in \mathbb{Z}^d} \left(\sum_{r=1}^d g^{-1}(|l_r|)^{1/2\xi} \right) |\gamma_l| < \infty$$

for ξ as in A.1, where g^{-1} is the inverse function of g given in A.2.

A.5 *K, k are continuous, real, even, integrable functions, and $\int K(u) du = 1$.*

A.6 *The lag window satisfies $\int \prod_{r=1}^d |u_r| |k(u)| du < \infty$.*

A.7 *The lag window satisfies $k(u) = 0$ for some $|u_r| > 1, r = 1, \dots, d$.*

A.8 *For some $q > 1$,*

$$\lim_{x \rightarrow 0} \frac{1 - k(x)}{\|x\|^q} = k_q$$

with some finite constant k_q .

Recall that if $M_n(S^0) = m_n S^0$ with m_n scalar and S^0 a symmetric positive definite matrix, then the mean-square error of kernel estimation is $O(m^{-2q} + m^d/n)$ uniformly in frequency, and the optimal rate of convergence is achieved by $m_n = n^{1/(2q+d)}$. Usually an optimal S is specified by some loss function and consistently estimated. For a stochastic matrix $\hat{M}_n = M_n(\hat{S}_n) = m_n \hat{S}_n$, where $\hat{S}_n \rightarrow_p S^0$ and m_n is deterministic. We prove the uniform consistency of kernel estimators \hat{f}^g and \tilde{f}^g based on \hat{M}_n .

Theorem 1 Assume A.1 to A.5 and that $\int |K(u)| \|u\| du < \infty$. Consider $\hat{M}_n = m_n \hat{S}_n$, where m_n is a deterministic sequence and $\hat{S}_n \rightarrow_p S^0$, S^0 symmetric positive definite. If $m_n^{-1} + m_n^d n^{-1/2} \rightarrow 0$, then

$$\|\tilde{f}^g - f\|_\infty \xrightarrow{p} 0.$$

If A.6 also holds, then

$$\|\hat{f}^g - \tilde{f}^g\|_\infty = O_p(m_n^d n^{-1}). \quad (5)$$

Next, we consider the asymptotic distribution of the process $(\tilde{f}^g - f)$ at arbitrary finite sets $\lambda_1, \dots, \lambda_Q \in \Pi^d$. Consider $M_n(S) = m_n S$, where m_n is a deterministic sequence, and define

$$\tilde{v}_n(\lambda, S) = (n m_n^{-d})^{1/2} (\tilde{f}_{M_n(S)}^g(\lambda) - E[\tilde{f}_{M_n(S)}^g(\lambda)]),$$

where $\tilde{f}_{M_n(S)}^g(\lambda)$ is the modified kernel estimator based on $M_n(S)$. We define $\hat{v}_n(\lambda, S)$ similarly (using $\hat{f}_{M_n(S)}^g$ instead of $\tilde{f}_{M_n(S)}^g$). Let \mathcal{N} be a compact set of symmetric positive definite matrices (\cdot) denotes the indicator function.

Theorem 2 Assume A.1 to A.5 and that $\int_{\Pi^d} |K(u)|^2 du < \infty$ and $M_n(S) = m_n S$ with a deterministic sequence m_n satisfying $m_n^{-1} + m_n^d n^{-1/2} \rightarrow 0$. Then, for any $Q \in \mathbb{N}$ and all finite sets $(\lambda_1, S_1), \dots, (\lambda_Q, S_Q)$ in $\Pi^d \times \mathcal{N}$, we have

$$(\tilde{v}_n(\lambda_1, S_1), \dots, \tilde{v}_n(\lambda_Q, S_Q))' \xrightarrow{d} (G(\lambda_1, S_1), \dots, G(\lambda_Q, S_Q))',$$

where $(G(\lambda_1, S_1), \dots, G(\lambda_Q, S_Q))$ has a Q -dimensional Gaussian with zero mean and covariance function

$$\begin{aligned} & \text{Cov}[G(\lambda_a, S_a), G(\lambda_b, S_b)] \\ &= (2\pi)^d (1 + \delta(\lambda)) f(\lambda)^2 \int_{\mathbb{R}^d} k(S_a^{-1}u) k(S_b^{-1}u) du \times I(\lambda_a = \lambda_b = \lambda) \end{aligned}$$

with $\delta(\lambda) = 1$ when the coordinates $\lambda_1, \dots, \lambda_d \in \{2\pi k : k \in \mathbb{Z}\}$ and $\delta(\lambda) = 0$ otherwise. If A.7 holds, the same result is satisfied by $\hat{v}_n(\lambda, S)$.

Instead of A.7, in the last statement of Theorem 2, we can use the condition A.6. The asymptotic distribution of $\hat{v}_n(\lambda, S)$ follows from (5) and the first part of Theorem 2.

Next, we ensure that the estimation of S^0 does not have an asymptotic effect on the limit distribution. Uniform weak convergence is proved applying from Bickel and Wichura (1971)

Theorem 3 *Under the conditions of Theorem 2, including A.7, assume that $E[|\varepsilon_t|^8] < \infty$ and that k is a Lipschitz function. Then*

1. *For any $Q \in \mathbb{N}$ and all $\lambda_1, \dots, \lambda_Q$, in Π^d , there exists a Gaussian process $G_Q(S)$ on $C(\mathcal{N})^Q$ such that*

$$(\tilde{v}_n(\lambda_1, S), \dots, \tilde{v}_n(\lambda_Q, S)) \xrightarrow{d} G_Q(S),$$

uniformly on $C(\mathcal{N})^Q$, where $G_Q(S)$ has zero mean and covariance function as in Theorem 2. If the conditions of Theorem 1 hold, then also

$$(\hat{v}_n(\lambda_1, S), \dots, \hat{v}_n(\lambda_Q, S)) \xrightarrow{d} G_Q(S),$$

uniformly on $C(\mathcal{N})^Q$.

2. *If, in addition, $\int |k(u)| \|u\|_1 du < \infty$ and $n^{-1/2} m_n^{d+1} = O(1)$, then, for any consistent estimator $\hat{S}_n \rightarrow_p S^0$, the process*

$$\tilde{v}_n(\lambda) = \tilde{v}_n(\lambda, \hat{S}_n) \xrightarrow{d} G^0(\lambda),$$

uniformly on $C(\Pi^d)$, where G^0 is Gaussian process with zero mean and covariances

$$\text{Cov}[G^0(\lambda_a), G^0(\lambda_b)] = (2\pi)^d (1 + \delta(\lambda)) f(\lambda)^2 k^2 \det(S^0) \times I(\lambda_a = \lambda_b = \lambda)$$

with $k^2 = \int_{\mathbb{R}^d} k(u)^2 du$, and, under the conditions of Theorem 1,

$$\hat{v}_n(\lambda) = \hat{v}_n(\lambda, \hat{S}_n) \xrightarrow{d} G^0(\lambda),$$

uniformly on $C(\Pi^d)$.

Let us consider

$$A_n(\lambda) = \frac{(n m_n^{-d})^{1/2}}{((2\pi)^d (1 + \delta(\lambda)) k^2 \det(S^0))^{1/2}} \left(\frac{\tilde{v}_n(\lambda, S)}{f(\lambda)} \right).$$

By Theorem 3 and the continuous mapping theorem, for all $\phi \in C(\Pi^d)$,

$$\int_{\Pi^d} \phi(\lambda) A_n(\lambda) du \xrightarrow{d} \int_{\Pi^d} \phi(\lambda) dW(\lambda),$$

i.e., the asymptotic distribution of $\int_{\Pi^d} \phi(\lambda) A_n(\lambda) du$ is $N(0, \|\phi\|_{L_2(\Pi^d)}^2)$. An interpretation for this behavior is that the asymptotic distribution of $A_n(\lambda)$ is that of \dot{W} , the Gaussian white-noise generalized process on $C(\Pi^d)$.

Note that Theorem 1 establishes the uniform consistency for kernel estimators when the smoothing number has been consistently estimated from the data. Theorem 3 establishes the weak convergence uniformly in $C(\Pi^d)$ when the smoothing number has been consistently estimated. Next, we consider the choice of the parameter S^0 , which is crucial to deal with the trade-off effects between the bias and variance. Let us define the stochastic process

$$\tilde{\alpha}_n(\lambda, S) = (n m_n^{-d})^{1/2} (\tilde{f}_{m_n S}^g(\lambda) - f(\lambda))$$

on $C^q(\Pi^d \times \mathcal{N})$. Under A.1 to A.5, A.8, and $M_n = m_n S$, applying an argument similar to that of Hannan (1970, Theorem 10, p. 283), we have that, if $f \in C^q(\Pi^d)$, then

$$m_n^q (E[\tilde{f}_{m_n S}^g(\lambda)] - f(\lambda)) \rightarrow \frac{k_q}{(2\pi)^d} S^{-q} \sum_l \|l\|^q \gamma_l e^{-il \cdot \lambda}, \quad (6)$$

uniformly in $\lambda \in \Pi^d$ and $S \in \mathcal{N}$, and therefore

$$\begin{aligned} \|E[\tilde{f}_{m_n S}^g] - f\|_{L_2}^2 &= \int_{\Pi^d} \left(\frac{k_q}{m_n^q (2\pi)^d} S^{-q} \sum_l \|l\|^q \gamma_l e^{-il \cdot \lambda} \right)^2 d\lambda + o(1) \\ &= \frac{k_q^2}{m_n^{2q}} S^{-2q} \sum_l \|l\|^{2q} |\gamma_l|^2 + o(1). \end{aligned}$$

Under the conditions of Theorem 3, A.8, and $f \in C^q(\Pi^d)$, the continuous mapping theorem implies that

$$\begin{aligned} &d \left(\int_{\Pi^d} |\tilde{\alpha}_n(\lambda, S)|^2 d\lambda, \left(\int_{\Pi^d} G(\lambda, S)^2 d\lambda + n m_n^{-(2q+d)} k_q^2 S^{-2q} \sum_l \|l\|^{2q} |\gamma_l|^2 \right) \right) \\ &\rightarrow 0, \end{aligned}$$

for any distance d that generates the weak-* topology on $C(\mathcal{N})$, where $E[\int_{\Pi^d} |G(\lambda, S)|^2 d\lambda] = \det(S) \kappa^2 \|f\|_{L_2}^2$. As a consequence, if we take $m_n = n^{1/(2q+d)}$, then

$$\int_{\Pi^d} |\tilde{\alpha}_n(\lambda, S)|^2 d\lambda \xrightarrow{d} \int_{\Pi^d} G(\lambda, S)^2 d\lambda + k_q^2 S^{-2q} \sum_l \|l\|^{2q} |\gamma_l|^2,$$

uniformly in $C(\mathcal{N})$. Therefore, we define the loss function

$$Q(S) = \det(S) \kappa^2 \|f\|_{L_2}^2 + k_q^2 S^{-2q} \sum_l \|l\|^{2q} |\gamma_l|^2$$

and define the optimal matrix S_0^* as a locally unique minimum for $Q(S)$. Similar arguments can be considered for $\hat{f}_{m_n S}^g$. The next section considers consistent plug-in and Bootstrap estimators of the optimal smoothing number S_0^* .

Finally note that, under the assumptions of Theorem 3 and A.8, if m_n satisfies the condition $n m_n^{-(2q+d)} \rightarrow 0$, then the asymptotic bias has the rate lower than $(n m_n^{-d})^{1/2}$, since

$$(n m_n^{-d})^{1/2} m_n^{-q} = (n m_n^{-(2q+d)})^{1/2} \rightarrow 0$$

and therefore $\tilde{\alpha}_n(\lambda) \rightarrow_d G^0(\lambda, \cdot)$, i.e., the asymptotic distribution of \tilde{f}^g after normalization concentrates around f without any asymptotic bias (see Hannan 1970, p. 288). Since the bias term tends to zero faster than the deviation term, we might consider the loss function given by the integrated variance $\det(S)\kappa^2 \|f\|_{L_2}^2$ and matrix S_0^* with the smallest determinant in the border of \mathcal{N} . Albeit for small samples, it is worthwhile to balance the bias and variance, e.g., by minimizing $Q(S)$.

4 Bootstrap and plug-in estimators

In this section, we consider the Bootstrap and plug-in estimations of S_0^* for the spectral estimator $\hat{f}_{m_n S}^g$, but similar arguments can be considered for $\hat{f}_{m_n S}^g$. The simplest approach is the plug-in estimation. Given a pilot estimator $\tilde{f}_{m_n \hat{S}_a}^g(\lambda)$, the plug-in loss function is defined by

$$Q_n^{\text{pi}}(S) = \det(S)\kappa^2 \|\tilde{f}_{m_n \hat{S}_a}^g\|_{L_2}^2 + k_q^2 S^{-2q} \sum_l \|\ell\|^{2q} |c_{n,l}^*|^2,$$

where $\sum_l \ell = \sum_{i=1, \dots, d} \sum_{|l_i| \leq g(n_i)}$. The plug-in estimator of S_0^* is given by the argument minimizing $Q_n^{\text{pi}}(S)$ on \mathcal{N} , i.e.,

$$\hat{S}_n^{\text{pi}} = \arg \min_{S \in \mathcal{N}} Q_n^{\text{pi}}(S).$$

Next, we define a bootstrap estimator of S_0^* . We consider a Wiener random field W_u^* on $C(\Pi^d)$, which is a multiparameter analogue of a Brownian motion with covariance function $\text{Cov}(W_u^*, W_v^*) = \prod_{r=1}^d \min\{u_r, v_r\}$, and define

$$\tilde{\alpha}_n^*(\lambda, S) = (n m_n^{-d})^{1/2} \sqrt{\frac{(2\pi)^d}{n}} \int_{\Pi^d} K_{m_n S}(u - \lambda) I_g(u) dW_u^*.$$

The conditional distribution of $\tilde{\alpha}_n^*(\lambda, S)$ with respect to the original sample is normal with mean $E^*[\tilde{\alpha}_n^*(\lambda, S)] = 0$ and variance

$$\text{Var}^*[\tilde{\alpha}_n^*(\lambda, S)] = m_n^{-d} (2\pi)^d \int_{\Pi^d} K_{m_n S}(u - \lambda)^2 I_g(u)^2 du.$$

The stochastic integral $\tilde{\alpha}_n^*(\lambda, S)$ is determinant in the Bootstrap method.

The evaluation of $\tilde{\alpha}_n^*(\lambda, S)$ requires the simulation of a continuous Wiener random field and computation of a multiparameter Itô integral, which is not feasible in practice, and discrete approximations are required. Thus, a discrete version can be considered, i.e.,

$$\tilde{\alpha}_n^{**}(\lambda, S) = (n m_n^{-d})^{1/2} \sqrt{\frac{(2\pi)^d}{n}} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda) I_g(\omega_{j,n}) W_{j,n}^*,$$

where $W_{j,n}^* = \prod_{r=1}^d (W^*(\omega_{j,n}) - W^*(\omega_{j-e_r,n}))$, and $I = (e_1, \dots, e_d)$ is the identity matrix. Note that $W_{j,n}^* = \prod_{r=1}^d \varepsilon_{n_r, j_r}$ with ε_{n_r, j_r} independently distributed $N(0, 2\pi j_r/n_r)$ for all $j \in J_n$. The expectation of $\tilde{\alpha}_n^{**}(\lambda, S)$ conditional to the sample is zero, and the variance is

$$\text{Var}^*[\tilde{\alpha}_n^{**}(\lambda, S)] = m_n^{-d} (2\pi)^d \left(n^{-1} \sum_{j \in J_n} K_M(\omega_{j,n} - \lambda)^2 I_g(\omega_{j,n})^2 \right).$$

The analogy with the multiparameter Itô integral is clear.

Next, the Bootstrap loss function is defined either in terms of the multiparameter Itô integral or using the discrete version, respectively, given by

$$Q_n^{b*}(S) = E^* \left[\int_{\Pi^d} |\tilde{\alpha}_n^*(\lambda, S)|^2 d\lambda \right] + k_q^2 S^{-2q} \sum_l \nu \|l\|^{2q} |c_{n,l}^*|^2,$$

$$Q_n^{b**}(S) = E^* \left[\int_{\Pi^d} |\tilde{\alpha}_n^{**}(\lambda, S)|^2 d\lambda \right] + k_q^2 S^{-2q} \sum_l \nu \|l\|^{2q} |c_{n,l}^*|^2,$$

where $E^*[\cdot]$ denotes the conditional expectation with respect to the data. In practice, the conditional expectation $E^*[\tilde{\alpha}_n^*(\lambda, S)^2]$ can be computed by Monte Carlo methods, e.g., by using the average of B realizations of $\tilde{\alpha}_n^*(\lambda, S)$, $B^{-1} \sum_{b=1}^B |\tilde{\alpha}_n^{*b}(\lambda, S)|^2$. Each of the values $\tilde{\alpha}_n^{*b}(\lambda, S)$ is computed by using an independent realization of the Brownian motion W_u^* . The discrete version $E^*[\tilde{\alpha}_n^{**}(\lambda, S)^2]$ can be similarly computed.

The Bootstrap estimator \hat{S}_n^{b*} of S_0^* is defined by the minimizer of $Q_n^{b*}(S)$, i.e.,

$$\hat{S}_n^{b*} = \arg \min_{S \in \mathcal{N}} Q_n^{b*}(S).$$

The Bootstrap estimator \hat{S}_n^{b**} can be defined similarly as the minimizer of $Q_n^{b**}(S)$ on \mathcal{N} .

The following result proves the consistency of the plug-in and Bootstrap estimators with respect to S_0^* .

Theorem 4 *Under the conditions of Theorem 3, if A.8 is satisfied, $f \in C^q(\Pi^d)$, and $m_n = n^{1/(2q+d)}$, then $\hat{S}_n^{\text{pi}} \rightarrow_p S_0^*$, $\hat{S}_n^{b*} \rightarrow_p S_0^*$, and $\hat{S}_n^{b**} \rightarrow_p S_0^*$.*

Under the conditions of Theorem 3, if \hat{S}_n is any consistent estimator for S_0^* , then both $\tilde{\alpha}_n^*(\lambda, \hat{S}_n)$ and $\tilde{\alpha}_n^{**}(\lambda, \hat{S}_n)$ have the same asymptotic distribution as that

of $(nm_n^{-d})^{1/2}(\tilde{f}_{m_n S_0^*}^g(\lambda) - E[\tilde{f}_{m_n S_0^*}^g(\lambda)])$.

Under the conditions of Theorem 3, A.8, and $f \in C^q(\Pi^d)$, we can obtain a bootstrap approximation to the distribution of $(nm_n^{-d})^{1/2}(\tilde{f}_{m_n S_0^*}^g(\lambda) - f(\lambda))$ by adding to $\tilde{\alpha}_n^*(\lambda, \hat{S}_n)$ a plug-in estimation of the asymptotic bias (6) scaled by $(nm_n^{-d})^{1/2}$; i.e., by considering

$$\tilde{\alpha}_n^*(\lambda, \hat{S}_n) + (nm_n^{-d})^{1/2} \left\{ -\frac{k_q}{m_n^q (2\pi)^d} \hat{S}_n^{-q} \sum_l \prime \prime \|l\|^q c_{n,l}^* \cos(l \cdot \lambda) \right\}.$$

A similar procedure can be used with $\tilde{\alpha}_n^{**}(\lambda, \hat{S}_n)$.

References

- Abramenko V, Yurchyshyn V, Wang H, Goode PR (2001) Magnetic power spectra derived from ground and space measurements of the solar magnetic fields *Sol Phys* 201:225–240
- Agterberg FP (1967) Computer techniques in geology *Earth Sci Rev* 3:47–77
- Barry RG, Perry A (1973) *Synoptic climatology* Methuen, London
- Beltrao KI, Bloomfield P (1987) Determining the bandwidth of a kernel spectrum estimate *J Time Ser Anal* 15:383–400
- Bickel PJ, Wichura MJ (1971) Convergence criteria for multiparameter stochastic processes and some applications *Ann Math Stat* 42:1656–1670
- Bolthausen E (1982) On the CLT for stationary mixing random fields *Ann Probab* 10:1047–1050
- Brillinger DR (1970) The frequency analysis of relations between stationary spatial series In: Pyke R (ed) *Proceedings of the 12th biennial seminar on time series, stochastic processes, convexity and combinatorics* Canadian Math Congress, Montreal, pp 39–81
- Brillinger DR (1981) *Time series, data analysis and theory* Holden Day, San Francisco
- Dalhaus R, Künsch H (1987) Edge effects and efficient parameter estimation for stationary random fields *Biometrika* 74:877–882
- Delgado M, Vidal-Sanz JM (2001) Averaged singular integral estimation as bias reduction technique *J Multivar Anal* 80:127–137
- Doukhan P (1994) *Mixing Properties and examples* Springer, New York
- Frank J, Härdle W (1992) On bootstrapping kernel spectral estimates *Ann Stat* 20:121–145
- Guyon X (1982) Parameter estimation for a stationary process on a d -dimensional lattice *Biometrika* 69:95–105
- Hannan EJ (1970) *Multiple time series* Wiley, New York
- Hannan EJ (1973) The asymptotic theory of linear time-series models *J Appl Probab* 10:130–145
- Heyde CC, Gay R (1993) Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence *Stoch Process Appl* 45:169–182
- Ivanov AV, Leonenko NN (1986) *Statistical analysis of random fields Mathematics and its applications, Soviet subseries* Kluwer Academic, Dordrecht
- Korovkin PP (1960) *Linear operators and approximation theory* Hindustan Publ Corp, Delhi
- Leonenko N (1999) Limit theorems for random fields with singular spectrum *Mathematics and its applications*, vol 465 Kluwer Academic, Dordrecht
- Politis DN, Romano JP (1996) On flat-top kernel spectral density estimators for homogeneous random fields *J Stat Plan Inference* 51:41–53
- Priestley MB (1964) The analysis of a two-dimensional stationary process with discontinuous spectra *Biometrika* 51:195–217
- Reshaw E (1984) Competition experiments in a plant monoculture: an analysis based on two dimensional spectra *Biometrics* 40:717–728
- Reshaw E, Ford ED (1983) The interpretation of process from pattern using two dimensional spectral analysis: methods and problems of interpretation *Appl Stat* 32:51–63
- Ripley BD (1988) *Statistical inference for spatial processes* An essay awarded the Adams Prize of the University of Cambridge Cambridge University Press, Cambridge

- Robinson PM (1991) Automatic frequency domain inference on semiparametric and nonparametric models *Econometrica* 59:1329–1363
- Robinson PM (2007) Nonparametric spectrum estimation for spatial data Special Issue in honour of Madam L. Puri *J Stat Plan Inference* 137:1024–1034
- Robinson PM, Vidal-Sanz JM (2006) Modified whittle estimation of multilateral models on a lattice *J Multivar Anal* 97:1090–1120
- Rosenblatt (1985) Stationary sequences and random fields Birkhäuser, Boston
- Rozañov YA (1967) Stationary random process Holden Day, San Francisco
- Rudin W (1974) Real and complex analysis, 2nd edn McGraw–Hill, New York
- Vidal-Sanz JM (2005) Pointwise universal consistency of nonparametric linear estimators *Bernoulli* 11(4):1–15
- Vidal-Sanz JM (2007) Automatic spectral density estimation for random fields on a lattice via bootstrap Working paper 07-26, business economics, Universidad Carlos III de Madrid, <http://e-archivo.uc3m.es:8080/dspace/bitstream/10016/680/1/wb072606.pdf>
- Whittle P (1954) On stationary processes in the plane *Biometrika* 41:434–449
- Zygmund A (1959) Trigonometric series, vols I and II combined, 2nd edn Cambridge mathematical library Cambridge University Press, Cambridge
- Žurbenko IG (1986) The spectral analysis of time series North Holland, Amsterdam