Spectral estimation for spatial point patterns

Tuomas A. Rajala[1], Sofia C. Olhede[2], and David J. Murrell[3]

[1] *tuomas.rajala@iki.fi, Applied statistical methods, Natural Resources Institute Finland (Luke), Helsinki, Finland*
[2] *École polytechnique fédérale de Lausanne, Lausanne, Switzerland, and Department of Statistical Science, UCL, London, UK*
[3] *Department of Genetics, Evolution and Environment, UCL, London, UK*

Abstract

This article determines how to implement spatial spectral analysis of point processes (in two dimensions or more), by establishing the moments of raw spectral summaries of point processes. We establish the first moments of raw direct spectral estimates such as the discrete Fourier transform of a point pattern. These have a number of surprising features that departs from the properties of raw spectral estimates of random fields and time series. As for random fields, the special case of isotropic processes warrants special attention, which we discuss. For time series and random fields white noise plays a special role, mirrored by the Poisson processes in the case of the point process. For random fields bilinear estimators are prevalent in spectral analysis. We discuss how to smooth any bilinear spectral estimator for a point process. We also determine how to taper this bilinear spectral estimator, how to calculate the periodogram, sample the wavenumbers and discuss the correlation of the periodogram. In parts this corresponds to recommending suitable separable as well as isotropic tapers in d dimensions. This, in aggregation, establishes the foundations for spectral analysis of point processes.

MSC 2020 subject classification: *62M30; 62M15*

Keywords and phrases: *Spectral density function; Spatial point processes; Debiased periodogram; Tapering*

1 Introduction

On one hand, point processes are a very important special form of stochastic processes. They have been the focus of numerous books, and are applied to model patterns in criminology, ecology and earth science, just to list a few concrete examples. On the other hand spectral/harmonic analysis of stochastic processes have led to making key insights into the nature of spatial and temporal processes. Despite the long development of spectral analysis in time series, hearkening back to the 1890's, their usage is far from standard in the spatial point process literature, in fact the lack of methodological development has virtually made application non-existent. This paper aims to develop the fundamentals of spectral analysis methodology for stationary point processes. Many of the basic choices for time series and random fields are not well developed for point processes, which we answer head on.

If one only observed a stochastic process once over a given domain, then additional assumptions are required to enable estimation, basically to enable averaging within rather than across ensembles. The standard assumption that enables averaging for estimation, is that of spatial homogeneity or stationarity. We shall also make this assumption throughout this article. Without this assumption any spectral decomposition needs additional care in interpretation, and as many questions remain outstanding even in this case, we confine our interest to stationary processes. In addition, it is possible to assume isotropy, so that dependence becomes characterised in one variable. We shall discuss isotropy, but not confine our interest to solely this special case.

The spectral decomposition of a random process on one hand gives a decomposition of variance of a stochastic process into a scale–by–scale contribution of energy; on the other hand, it gives the representation of the process as a weighted average of complex exponentials so that given behaviour can be associated with a given physical scale (the inverse of its wavenumber or magnitude wavenumber). The latter notion of scale is key to many further developments in the representation of a particular stochastic process. For point processes the corresponding results are harder to derive as our observations do not take Euclidean values at fixed regular locations but are either absent or present at that given (but random) location. This means that the random location enters non-linearly in the Fourier transform, as the argument of the complex exponential, rather than as in a linear combination.

Spectral analysis of regularly sampled processes has already been studied extensively in the 1-D case (Percival and Walden, 1993b), but even as we study results into 2-D understanding the corresponding developments

become more challenging, especially as non-parametric methods can become hard to interpret. For random fields (Stein, 2012; Adler and Taylor, 2010; Rosenblatt, 2012; Deb et al., 2017; Guillaumin et al., 2019) spectral analysis does resemble the analysis of time series, but the additional challenges of embedded simplified structure (lower dimensional structure) in 2-D and higher D are harder to address, and the appropriate visualization remains an outstanding challenge, see for example (Errico, 1985; Durran et al., 2017; Osgood, 2019). Also, there is some study of random fields sampled by 2-D point processes, as for example by (Chou and Lii, 1992), in 1-D (Lii and Masry, 1994).

Despite the potential promise of spectral analysis of point processes, after the initial flurry of development (Bartlett, 1963, 1964; Mugglestone and Renshaw, 1996b), also summarized in (Daley and Vere-Jones, 2003, Ch. 8), some results followed for its practical utility (Mugglestone, 1990; Mugglestone and Renshaw, 1996b, 1998), but its promise was never realised into broad usage. Somehow spectral analysis was never converted into a well–used practical tool set for data analysis, and normally it is referred to only when modelling and analysis depend on it (Lavancier et al., 2015; Møller and Torrisi, 2007). We suspect the reason why is, at least partly, that a number of practical choices necessary for constructing spectral summaries have never been substantiated by the proponents of spectral analysis from a theoretical or methodological point-of-view, and therefore remain somewhat mysterious and unclear. Naturally in the limit of low wavenumbers spectral analysis of point processes must coincide what with spectral analysis of random fields to be self–consistent, but despite this fact, even at low wavenumbers, there are many unexpected features as we shall show. We shall assume that the estimation of the spectrum is the end point of the analysis of the point process.

We can simply contrast this state-of-the-art with the results available for random fields (Rosenblatt, 2012; Adler and Taylor, 2010; Stein, 2012). The usage of spectral methods for point processes is limited in parts just because the Cooley–Tukey FFT algorithm (Cooley and Tukey, 1965) cannot be used when performing analysis when by definition, the spacing of observations is not regularly spaced.

The aim of this paper is to understand the properties of any direct spectral estimator defined from a spatial point process, assuming we observe a spatially stationary process. The spectral density captures all relevant second order features of the process, and so this is a fundamental characterization of the process in question. Taking inspiration from standard practice in signal processing, we shall focus on the class of bilinear spectral estimators (Cohen, 1995) that form a quadratic either in the Direct Fourier Transform (DFT) of the data or the data directly, or a linearly modified version, bias-correcting or such as using tapering (Walden, 2000).

The well-known periodogram non-parametric spectral estimator is naturally a special (and important) case of a bilinear spectral estimator. Two important tools from time series, of de-trending (Percival and Walden, 1993b; Velasco and Robinson, 2000) and tapering (Thomson, 1982) will be transferred to point processes, and they will have a unique form in this context. It goes without saying, that the point process setting is much more involved than that of time series, and we introduce appropriate separable and radial tapers, drawing on our expertise from signal processing.

To understand the spectral analysis of point processes therefore we will address the following key points:

1. How do we define spectral summaries for a point process? We recall existing frameworks for spectral analysis of a point process in Section 2.

2. We discuss in Section 3 how direct spectral summaries (Percival and Walden, 1993b) are defined, both for general point processes and isotropic processes, drawing upon understanding from spectral analysis for random fields. This includes understanding how to taper a spatial point process.

3. To define method of moments estimators, we answer the question: What are the moments of the DFT? This is a more complex question than that for random fields or time series, and is treated exhaustively in Section 4.

4. To understand linear estimation we need both to determine the bias and variance/covariance; Section 5 determines the second order properties of direct spectral estimators.

5. Armed with the bias and variance, we can address non-parametric estimation. This question is treated in Section 6 and Section 8.

6. How does tapering of a point process work? This is further discussed in Section 6.

7. How should we compute the periodogram (the spectral "energy" of a process at a given wavenumber), and does it matter if we assume the process is isotropic or anisotropic? This question is further treated in Section 7.

8. How do we sample the spectrum in wavenumber? Both to achieve near uncorrelatedness, but also determining when the quality of the estimator becomes so poor that the periodogram contains no more information.

All these points will be clarified in order to enable spectral estimation, and determine the mean square error of a spectral estimator.

Note that the Direct Fourier Transform (DFT) of the point process is not zero–mean under the assumption of stationarity, unlike for random fields, as has been noted by (Diggle et al., 1987), and discussed more abstractly in terms of random measures by (Daley and Vere-Jones, 2003). We refer to work by (Biscio and Waagepetersen, 2019) and deduce Gaussianity of the DFT. We show that the DFT of a point process will not be zero–mean, no matter its marginal distribution (compare with discussion in Diggle et al. 1987 for small wavenumbers). To be concrete, we shall determine the first and second moments of the DFT, which, combined with the Gaussian limiting results, will completely specify the distribution (under growing–domain asymptotics, compare with Stein (1995) for fixed-domain asymptotics).

The Fourier Transform (FT) will be (modulus) squared to yield the periodogram, which will have asymptotically a non-central chi-square distribution, but after de-trending, this will be a chi-square just like the modulus square of the DFT of a random field. This shows that we cannot simply assume that the properties of DFT of a random field will be directly inherited by a FT of a point process, and shows the challenges in using Diggle's (Diggle et al., 1987) isotropic estimator, as the problem of its sampling in wavenumber is much harder to treat.

Tapering (Walden, 2000) is the second method that we shall explore more fulsomely. It is not clear how to taper in the point process context, or what taper to use, or indeed how to use them. Tapers are normally defined for regularly spaced data. We shall use more rarely applied continuous–space tapers, and show for what classes of processes, multi–tapering produces reduced variance estimators. We note that other choices of implementing tapering could have been made; and this will be discussed, such as tapers for irregularly spaced stochastic processes. We shall finish up with an extensive simulation study to illustrate the performance of our proposed estimators.

Therefore, we aim to settle many basic outstanding issues in spectral analysis of point processes. We also connect to related work on sampling a continuously indexed random field, see (Chou and Lii, 1992) for example (already developed fully for 1-D processes by Lii and Masry 1994). Our end result is the arrival at an appropriate smoothing method for the raw spectral estimator of a point process, and also suitable 1-D summaries of such a process, whether the process is isotropic or not. Section 7 discusses isotropic processes and isotropic summaries. This all combines to underpin our understanding of the spectral representation of point processes. Section 8 explores the implications of our results via simulations. We conclude the paper in Section 9 with a discussion of the derived results.

2 Notation and Definitions

We assume that we observe a spatially homogeneous point process X with intensity $\lambda = \rho^{(1)} > 0$ and with $\rho^{(2)}(z)$ defined as the (second order) product density of X. This means that $\rho^{(2)}(z)$ only has one argument ($z \in \mathbb{R}^d$, namely the spatial shift) rather than depending on two local spatial variables (say x and y rather than $z = x - y$). We assume that X is a simple point process on $\mathbb{R}^d$, meaning no duplicate points are allowed. We take $d = 2, 3, \ldots$, and so exclude the case of $d = 1$, which is relatively well-studied, see references in e.g. (Daley, 1971; Jowett et al., 1972; Vere-Jones, 1974). For a (Borel) set A in $\mathbb{R}^d$, we define $|A|$ for the volume of A and $N(A)$ for the number of points of $X \subset \mathbb{R}^d$ in A. We denote by B any subset of $\mathbb{R}^d$ where points are observed (or equivalently our observation domain).

In complete analogue with a random field we shall define the spectral density of X as the Fourier transform of the complete covariance function of X. This is the standard approach, and was first proposed by Bartlett (1964). The complete covariance function of a stationary point process is therefore

$$\gamma(z) \equiv \lambda\delta(z) + \rho^{(2)}(z) - \lambda^2, \quad z \in \mathbb{R}^d, \tag{1}$$

where $\delta(\cdot)$ is Dirac's delta function. The spectral density function (sdf) (Bartlett, 1964) of the point process X is then defined as the Fourier transform of the complete covariance function:

$$f(w) \equiv \mathcal{F}[\gamma](w) \; = \int_{\mathbb{R}^d} e^{-2\pi i w\cdot z}\gamma(z)\,dz \; = \; \lambda + \int_{\mathbb{R}^d} e^{-2\pi i w\cdot z}[\rho^{(2)}(z) - \lambda^2]dz, \qquad w \in \mathbb{R}^d. \tag{2}$$

This representation is in direct analogy with the corresponding spectral decomposition of a random field or a time series. The symbol $\mathcal{F}$ denotes the Fourier transform, with i as the imaginary unit. We refer to the argument w as the "wave number" rather than frequency to knowledge its multi-dimensionality. This is a natural choice of a Fourier transform in analogy to the analysis of random fields (Stein, 2012). To avoid additional constants in the inverse Fourier transforms, we parameterise the Fourier transform with wave number instead of the customary angular frequency $\xi = 2\pi w$ used by Bartlett (1964) and some others.

For a time series or random field there are a number of spectral results, ranging from the spectral representation theorem (I) (Percival and Walden, 1993a, pp.130) and to Bochner's (Herglotz) theorem (II) (Brockwell

and Davis, 1991; Percival and Walden, 1993a). Both these results are not exactly available in the point process setting, but for a time series X_t we can note them:

$$X_t \overset{(I)}{=} \mu + \int_{-\frac{1}{2}}^{\frac{1}{2}} dZ(\omega) e^{2i\pi\omega t}, \quad \mathbf{Var}\{dZ(\omega)\} \overset{(II)}{=} f(\omega) d\omega \tag{3}$$

$$\gamma(t) \overset{(III)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\omega) e^{2\pi i\omega t}\, d\omega, \quad f(\omega) \overset{(IV)}{=} \sum_{\tau} \gamma(\tau) e^{-2i\pi\omega\cdot\tau}. \tag{4}$$

Equation (3) decomposes X_t into random contributions associated with each angular frequency ω. The covariance in Eqn (4) is also decomposed into weighted frequency contribution. Both of these observations are important for the interpretation of the Fourier representation of X_t. "Important" contributions correspond to $f(\omega)$ being considerably larger relative to other $f(\omega)$ as in that scenario $\mathbf{Var}\{dZ(\omega)\}$ is bigger than $\mathbf{Var}\{dZ(\omega')\}$ and therefore $|dZ(\omega)|$ likelier to be larger than $|dZ(\omega')|$. Spectral representations of point processes are also discussed in (Daley and Vere-Jones, 2003, Ch. 8). We can yet again decompose the (complete) covariance in terms of a spectral representation like in (II), and it takes the form in terms of a d-dimensional Fourier transform of

$$f(w) = \int e^{-2\pi i w\cdot z}\gamma(z)\, dz = \lambda + \int e^{-2\pi i w\cdot z}\left\{\rho^{(2)}(z) - \lambda^2\right\}\, dz. \tag{5}$$

We refer to $f(w)$ as the spectrum of X. We note that the Fourier transform can be inverted to yield the equality of (an analogy of III):

$$\rho^{(2)}(z) = \lambda^2 + \int_{\mathbb{R}^d} f(w) e^{2\pi i w\cdot z}\, dz. \tag{6}$$

We assume the spectrum is the primary object of interest in our study of point patterns, as the covariance $\gamma(z)$ can be fully determined from it, and the covariance specifies the point process. We see directly from (5) that $\rho^{(2)}(z) - \lambda^2$ and *not* the complete covariance $\gamma(z)$ is playing the role of a time series covariance. This is as when transforming $\gamma(z)$ rather than $\rho^{(2)}(z) - \lambda^2$ we otherwise get a constant boost of λ that spreads across all frequencies. This does not indicate to us which frequencies are "more" important unlike for a random field. By important we here mean contribute to the process more than any other wavenumber.

We have to exercise some caution when interpreting the Fourier transform as a bijective transform. Yes, it should contain the same information about scale, but the meaning of the word scale will be different. In time or space the notion is associated with the support of $\gamma(z)$ or $\rho^{(2)}(z)$. In frequency being supported over a wavenumber w means variation over scales $1/\|w\|$ is present in the correlation function, or to represent the variability in $\gamma(z)$ we need frequencies $\|w\|$. The function $\rho^{(2)}(z) - \lambda^2$ is often approached in terms of what scales it is non-zero at, but the variation in the function can be associated with long or short scales. So for a covariance function, we now have two notions of what it means to possess scales $\|w\|$; either $\gamma(1/\|w\|\mathbf{u})$ is non-zero for unit vector $\mathbf{u}$ or $\gamma(z)$ is variable (changing) and the scale of the change is $\|w\|$. Imagine observing a sinusoid with period k. The function will hit unity at a regular set of z values, and so it will be supported at those z values, but also its Fourier transform will be supported at $1/k$.

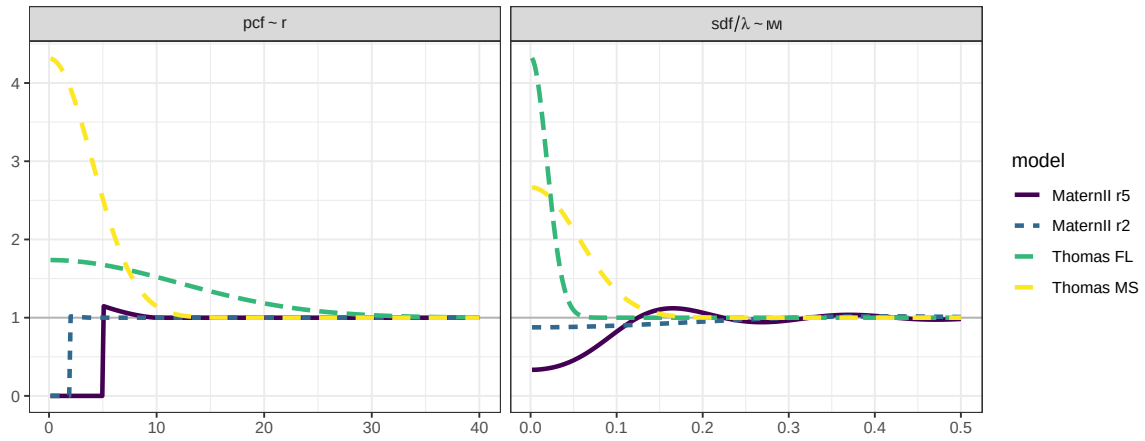

Figure 1: Pair correlation $pcf = \rho^{(2)}/\lambda^2$ (pcf, left) and corresponding scaled spectral density function f/λ (sdf/λ, right) for two stationary and isotropic non-Poisson point process models, two variants each, with known pcfs. For the Thomas processes the sdf is given in the text. For the regular Matern II process the sdf was numerically approximated using the Hankell transform (cf. Section 7). See details of the processes and the variants in Section 8.

This is clear from Figure 1 where covariance is present at smaller or medium scales smoothly for the clustered Thomas processes, or repelled for the Matérn hard-core processes. As the Matérn process's complete covariance function is discontinuous, its Fourier transform is supported over all scales (due to the Heisenberg-Gabor uncertainty principle (Cohen, 1995)). This can be deduced from (6). To reproduce the discontinuity we need all scales.

We see from (5) that not all frequencies or wavenumbers are given equal weighting. First all wavenumbers are given an equal weighting λ and then the Fourier transform of $\rho^{(2)}(z) - \lambda^2$ determines the wavenumbers that are up–weighted (added) or down–weighted (subtracted) relative to the overall level of λ. We then observe what wavenumbers are important to the representation of the point process, which gives more information, conveniently decomposed on a scale–by–scale manner.

For a random field or a stochastic process in d–dimensions a few cartoon characteristics are important. For a discretely sampled process in $\mathbb{Z}^d$ the simplest random process, white noise, is constant across wavenumbers yielding a spectrum that takes the form of σ^2 on $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]^d$, where $\Delta > 0$ is the sampling period, and zero otherwise. For a point process the choice of definition of the spectral density does not imply a decay because of the inclusion of the term $\lambda\delta(z)$ in space. However once we remove this term, we expect a decay of the remaining spectrum $f(w) - \lambda$ as $|w| \to \infty$. Otherwise (5) would contain additional singularities.

Furthermore at $w = 0$ we retain

$$f(0) = \lambda + \int \left\{ \rho^{(2)}(z) - \lambda^2 \right\} \, dz. \tag{7}$$

The second term in this expansion is not required to be zero, and for the Thomas process, it is not zero. For a Thomas process the spectrum is (Mugglestone and Renshaw, 1996b, p 55)

$$f(\omega) = \lambda_0 + \lambda_0' e^{-\frac{1}{2}a^2 \omega \cdot \omega}. \tag{8}$$

It is clear from this expression that we cannot arrive at a constant contribution due to the exponential.

As we shall be studying the moments of a point process, it is convenient to restate Campbell's theorem (Chiu et al., 2013, Sec. 4.3.3), and we shall use this result multiple times. The theorem applies to any measurable function $h : R^{nd} \mapsto R^+$ with $n = 1, 2, ...$, and states that

$$\mathbf{E} \sum_{x_1,...,x_n \in X}^{\neq} h(x_1, ..., x_n) = \int_{\mathbb{R}^{nd}} h(x_1, ..., x_n) \rho^{(n)}(x_1, ..., x_n) dx_1 ... dx_n, \tag{9}$$

where the summation is over distinct point tuples. Note that if the point pattern X is stationary then $\rho^{(1)}(x) = \lambda$ is constant for any x, and $\rho^{(n)}(x_1, ..., x_n) = \rho^{(n)}(o, x_2 - x_1, ..., x_n - x_1) \equiv \bar{\rho}^{(n)}(x_2 - x_1, ..., x_n - x_1)$. We shall use the $\bar{\rho}^{(n)}$ notation to explicitly avoid confusion when parameterizing in terms of the shift $z_1 = x_2 - x_1, \ldots, z_{n-1} = x_n - x_1$ (and the stationary parameterization) rather than the basic coordinates.

We define a new parameterisation to capture how the process's $f(w)$ differs from that of a Poisson process. We define the nth deviation as

$$\widetilde{\rho}^{(n)}(z) = \frac{\bar{\rho}^{(n)}(z) - \lambda^n}{\lambda^n}, \quad n = 2, 3, \ldots. \tag{10}$$

For a Poisson process these $n = 2, 3, \ldots$ deviations from Poissonianity will all be identically zero. By understanding deviations from Poisson behaviour, we get greater understanding of the underlying process under study. For completeness we also define the Fourier transform of the deviations:

$$\widetilde{f}^{(n)}(w) = \mathcal{F}\left\{ \widetilde{\rho}^{(n)}(z) \right\},$$

where if $n = 2$ we suppress the superscript. With this definition we find that

$$\widetilde{f}(w) = \frac{f(w) - \lambda}{\lambda^2}, \quad w \in \mathbb{R}^d, \quad f(w) = \lambda + \lambda^2 \widetilde{f}(w). \tag{11}$$

Thus in a sense, $\widetilde{f}(w)$ encapsulates the deviation of the function from a constant spectral density of λ via the term $\lambda^2 \widetilde{f}(w)$.

Daley and Vere-Jones (2003, p. 337) discusses some differences in spectral representation of time series versus that of point processes. We would argue that decay of the spectrum is still reasonable to assume once λ has been subtracted (so the decay of $\widetilde{f}(w)$ is reasonable to assume). Having established the theoretical spectral description of point processes, we now turn to their sampling properties.

3 Direct Spectral Summaries of the Data

We shall re-examine spectral estimation of point processes, and clarify a number of points that are currently unclear. We start from Bartlett's periodogram estimator based on observing a point pattern in region $B \subset \mathbb{R}^d$ which can be written as:

$$I_0(w) := \hat{\lambda} + |B|^{-1} \sum_{x \neq y, \in X \cap B} e^{-2\pi i w \cdot (x-y)}, \quad w \in \mathbb{R}^d, \tag{12}$$

where we have set $\hat{\lambda} = |B|^{-1}|X \cap B|$. This definition uses all the data available to us, $\{x : x \in X \cap B\}$, but is simply *a possible choice* amongst all possible direct spectral estimators. In general a direct spectral estimator is one which is bilinear in the DFT of the data, see for example the discussion in (Percival and Walden, 1993a, Ch 5–6). We recall a bilinear form is a function $D(\mathbf{x}, \mathbf{y})$ satisfying $D(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = D(\mathbf{x}_1, \mathbf{y}) + D(\mathbf{x}_2, \mathbf{y})$, and equivalently in the second argument.

Bilinear forms have been thoroughly discussed in signal processing for the analysis of time series (Loughlin et al., 1993). As a point process consists of locations the best we can hope for in terms of bilinearity will be in terms of sesquilinearity of the Fourier transform as the point locations will appear in the argument of the complex exponential. Sesquilinearity of $D(\mathbf{x}, \mathbf{y})$ simply generalizes a bilinear form to the Hermitian symmetry, additionally requiring $D(\mathbf{x}, \mathbf{y}) = D(\mathbf{y}^*, \mathbf{x})$ as well as $D(\mathbf{x}, \mathbf{y})$ satisfying $D(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = D(\mathbf{x}_1, \mathbf{y}) + D(\mathbf{x}_2, \mathbf{y})$.

In time series the bilinear form has been chosen to ensure that spectral estimators are real-valued, and often non-negative though some bilinear estimators are not, see e.g. the Kirkwood-Rihaczek spectrum (Hindberg et al., 2006), and Guyon's spectral estimator (Guyon, 1982). Spectral estimators in time series are bilinear in the observed real-valued process X and sesquilinear in its Fourier transform $J_h(w)$. We shall still require that the form of the estimator is sesquilinear in the DFT of the point process, but the estimators will not be bilinear in X, the point pattern, as this will not be possible. We define the tapered DFT of X of a point pattern for a specified general (square) integrable function $h(x)$ (referred to as a 'data taper' by Percival and Walden 1993b) with Fourier transform $H(w)$, to be

$$J_h(w) := \sum_{x \in X \cap B} h(x) e^{-2\pi i w \cdot x}, \quad w \in \mathbb{R}^d, \tag{13}$$

for any unit norm h, e.g. $\|h\|_2 = 1$.

Because x appears in the argument of the complex exponential, we cannot define estimator bilinear in the data. In practice one has to decide what taper function to use. In 1D for point processes tapering has been used (Cohen, 2014; Taleb and Cohen, 2018) with continuous wavelets, but also multitapering has been used on interpolated data (Das and Babadi, 2020). Here, we do not interpolate, do not implement localized analysis, and do not implement analysis in 1D. Higher dimension is more complex to interpolate, as it is not orderly.

We note that in general $h(x) \exp(-2\pi i w \cdot x)$ is identically distributed but not independent for many choices of point processes x, making the choice of a Central Limit Theorem (CLT) a bit more complex. For a large class of processes one such CLT is given by Biscio and Waagepetersen (2019). If we compare the quantity in (13) with (Biscio and Waagepetersen, 2019), then we see that $q = 1$ and $p = 2$, in their notation. We do not have to worry about $e^{-2\pi i w \cdot x}$ being a function of several of the points, and this is why $q = 1$. Requirement $\mathcal{H}1$ is therefore trivially not a problem as an assumption, $\mathcal{H}2$ is about the scaling of the observation area, as we assume the sides scale the same not a problem, $\mathcal{H}3$ (higher order centred powers) are limited through the usage of Campbell's theorem and as $q = 1$ we do not have to worry about pathological scaling between components in $\mathcal{H}4$. Therefore, citing (Biscio and Waagepetersen, 2019), we can deduce from their Theorem 1 that as $|B|$ diverges, and recover the second moments by direct calculations in Section 4:

$$J_h(w) \xrightarrow{L} N_C\left(\lambda H(w), f(w), r(w)\right), \tag{14}$$

$$r(w) = \lambda \int_{\mathbb{R}^d} H(w' - 2w) H(w') dw' + \int_{\mathbb{R}^d} U(w, z) e^{-2\pi i w \cdot z} \{\rho(z) - \lambda^2\} dx dz. \tag{15}$$

Here $|B| \to \infty$ in equal speed in all directions of B and B is the cuboid, and we define the tapered related term $H(w)$ in (21). We deduce the moments from Proposition 5.1 combined with Theorem 5.1. We can view $J_h(w)$ as a complex-valued scalar or a real-valued two-vector. $N_C(\mu, \Sigma, C)$ is the general complex normal (Schreier and Scharf, 2003) and its arguments are its mean μ, it covariance Σ, as well as its complimentary covariance C. It should be contrasted with the complex propert normal $N_C(\mu, \Sigma)$, that has zero relation.

Based on $J_h(w)$ the spectral estimator becomes its modulus square or

$$I_h(w) := |J_h(w)|^2 = \left| \sum_{x \in X \cap B} h(x) e^{-2\pi i w \cdot x} \right|^2 = \sum_{x \in X \cap B} h(x) h^*(x) + \sum_{x, y \in X \cap B}^{\neq} h(x) h^*(y) e^{-2\pi i w \cdot (x-y)}. \tag{16}$$

If we take $h(z) = h_0(z) = \frac{1}{\sqrt{|B|}} 1(z \in B)$ then we recover Bartlett's peridogram in (12) from (16). The point of using a general function $h(z)$ is that abruptly ending the inclusion of points when we leave the region B causes ripples in wavenumber, and therefore a worse estimation of the spectrum, as is also the case of time series when not using a taper (Percival and Walden, 1993b). We have deduced an asymptotic distribution for $J_h(w)$ using (Biscio and Waagepetersen, 2019) but this result cannot be applied to $I_h(w)$ even if it superficially may seem to be of the correct form.

What can we then say about $I_h(w)$? Using the continuous mapping theorem (DasGupta, 2008) we can deduce from (14) that if we consider only arguments w so that the complementary covariance is negligible then

$$J_h(w) \xrightarrow{L} N(\lambda H(w), f(w)) \Rightarrow \quad |J_h(w)|^2 \xrightarrow{L} \frac{f(w)}{2} \chi^2_2(\lambda^2 |H(w)|^2), \tag{17}$$

where the parameters of the non–central $\chi^2_k(m^2)$ (k denoting the degrees of freedom, and m^2 the non-centrality parameter). We give $H(w)$ in (21). We can also use the uniform integrability of the variable $J_h(w)$, to get the asymptotic moments of $I_0(w)$ from these limits.

Note that the sum in (16) cannot be evaluated in a computationally efficient manner (compare to (Cooley and Tukey, 1965)), as the locations $\{x\}$ are not regularly spaced, which is unfortunate, but unavoidable. Finally (16) is bilinear in $J_h(w)$ but it is not bilinear in X, unlike the commensurate expressions for the DFT of time series and random fields.

For completeness we also note the isotropic estimator of (Diggle et al., 1987, Eqn. 3.3) in 2D, namely

$$\bar{I}_0(\|w\|) := \hat{\lambda} + \frac{1}{|B|} \sum_{x,y \in X \cap B}^{\neq} \mathcal{J}_0\left(2\pi \|w\| \|x - y\|\right), \tag{18}$$

where $\mathcal{J}_0(x)$ denotes the Bessel function of order 0, specified in Section 7. The d-dimensional extension is given later in equation (24). $\bar{I}_0$ is less clearly bilinear in the data, but if we start from the estimator of (12) and average it over orientations analytically, then we arrive at this form. We therefore abuse our terminology to also refer to it as a bilinear estimator. There is one additional correction made by (Diggle et al., 1987). As we shall see, for low wave numbers there is a bias inherent in (18). To address this problem (Diggle et al., 1987, Eqn. 3.4) suggests taking for some choice of lower bound t_0

$$I_D(\|w\|) = \begin{cases} \bar{I}_0(t_0) & \text{if} \quad \|w\| \leq t_0 \\ \bar{I}_0(\|w\|) & \text{if} \quad \|w\| > t_0 \end{cases}. \tag{19}$$

Diggle et al. (1987) suggest taking t_0 so that $\bar{I}_0(t_0)$ is at a minimum, and also suggest smoothing the estimated I_D, which we clarify in the linear estimation section. The authors also discuss iterated methods of bias correction in their Section 5, and correcting biased estimators of the correlation. They discuss difficulties of preserving short range features in the spectrum.

Note also that we have modified the estimator in (18) to divide by $|B|$ rather then the observed number of points in the region, $N(B)$, as do for example (Diggle et al., 1987), as it is much preferable not to have a random denominator. We shall discuss the usage of $I_0(\cdot)$ versus $I_D(\cdot)$ further on in Section 7, and the implications of when the process is isotropic or not.

As can be surmised from (16) the expectation of the general estimator $\mathbf{E}\{I_h(w)\}$ is a convolution between the Fourier transform of the window and the true spectrum, and if $h(z)$ does not go to zero nicely at the boundary of B then the periodogram becomes quite complicated in terms of its expectation. Looking at (16) also raises a second problem, namely that there is a constant contribution $|B|^{-1}|X \cap B|$ which does not give any wave number specific information, and is correlated between wave numbers. This term is new to time series and random fields, if not new to the expectation of the periodogram of randomly sampled stochastic processes, see (Lii and Masry, 1994; Chou and Lii, 1992).

We can generalize the family of estimators from Eqn. (16) to take the form of

$$I_h(w) := \left| \sum_{x \in X \cap B} h(x) e^{-2\pi i w \cdot x} \right|^2 = |J_h(w)|^2, \quad w \in \mathbb{R}^d. \tag{20}$$

We study the *family of estimators* $I_h(w)$. We also note that (20) is a direct spectral estimator (Percival and Walden, 1993b, p. 207). Multidimensional tapers are available to us (Hanssen, 1997; Simons and Wang, 2011), and can be precomputed. Our choice of tapering corresponds to using continuous tapers evaluated at random locations.

Should we use no data taper, or equivalently use the classical taper h_0 from (8), then we introduce the notation of writing $I_0(w)$ for the periodogram for clarity. We then also define

$$H = \mathcal{F}[h], \quad H_0 = \mathcal{F}[h_0]. \tag{21}$$

The expectation of any member of that family takes the form of:

$$\mathbf{E}\left\{I_h(w)\right\} = \int_{\mathbb{R}^d} |H(w'-w)|^2 f(w')dw' + \lambda^2 |H(w)|^2. \tag{22}$$

This can be derived straightforwardly by direct computation from (20) by taking expectations with Campbell's formula and using the convolution theorem.

Should we chose to taper isotropically in 2-D-space then instead we take with isotropic taper $h_I(\cdot)$

$$\bar{I}_h(\|w\|) := \hat{\lambda} + \sum_{x,y \in X \cap B}^{\neq} h_I\left(\|x-y\|\right) \mathcal{J}_0\left(2\pi\|w\| \cdot \|x-y\|\right). \tag{23}$$

Note that this cannot necessarily be constructed. Say $J_h()$ with an isotropic taper would be made with the largest spherical domain and a compact support could be constructed, but what does that imply for $I_h()$? Radial tapers can be determined as described in (Simons and Wang, 2011), whether in 2-D continuous space or discrete space. Equation (23) can be of course be extended into higher dimensions. In general we would have in dimension $d = 1, 2, 3, 4, \ldots$

$$\bar{I}_h(\|w\|) := \hat{\lambda} + \sum_{x,y \in X \cap B}^{\neq} h_I\left(\|x-y\|\right) \mathcal{J}_{d/2-1}\left(\|w\|\|x-y\|\right) \|x-y\|^{-(d/2-1)}. \tag{24}$$

We can see directly that

$$\mathbf{E}\{\bar{I}_h(\|w\|)\} = \lambda + \int_B \int_{B \cap B_{-z}} h_I\left(\|z\|\right)\ \mathcal{J}_{d/2-1}\left(2\pi\|w\|\|z\|\right)\ \|z\|^{d/2-1}\ \tilde{\rho}^{(2)}(\|z\|)\ dzdx.$$

As the change of variable to polar coordinates means we will multiply by $\|z\|^{d-1}$, and this will correspond to a $d/2-1$ Hankel transform of the taper $h(\cdot)$ multiplied by the product density. For more details on the Hankel transform see (Grafakos and Teschl, 2013; Cressie, 2015).

We can also use more than one taper, and will use a sequence of orthogonal functions $\{h_k\}$ (Percival and Walden, 1993b), that will be used to get a new estimator. We will assume they are all of unit energy and are all (pairwise) orthogonal. This means we will require $||h_k||_2 = 1$ and

$$\int_{\mathbb{R}^d} h_k(z) h_l(z)\, dz = \delta_{kl}. \tag{25}$$

Tapering has been viewed with variable approbation. Criticism comes especially for throwing away data (Fougere, 1977; Yuen, 1979). We shall study the properties of direct spectral estimators of the form of Eqn (20). This will help us to characterise the second order properties of the process X.

Just as a brief note: it is not obvious to know how to taper. Most of time series analysis is based on discrete regular sampling for instance. We chose to use continuous tapers, rather than interpolating the points to a regular grid. This leaves the spheroidal wavefunctions (continuous but hard to compute), as well as the cosine tapers of Riedel and Sidorenko (1995). These correspond to separable taper choices; non-separable choices with be discussed in Section 7.

4 Expectation of Bilinear Spectral Estimators

We shall start from the simplest spectral estimator, namely the Bartlett periodogram as specified by (12). We shall now be quite concrete and understand some common cases of spatial data, and their sampling, and choose as early special cases Cartesian product domains. We shall focus on the cuboid sampling domain:

$$B_\square = [0, \ell_1] \times \cdots \times [0, \ell_d].$$

Lemma 4.1. *Assume that X is a homogeneous point process with intensity λ and twice differentiable spectrum $f(w)$ at all values of $w \neq 0$. Assume X is observed in a cuboid domain $B_\square$ with a centroid at 0, which is growing in every dimension. Then the expected value of the periodogram $I_0(w)$ satisfies the equation*

$$\mathbf{E}I_0(w) = f(w) + |B|^{-1}\lambda^2 T(B,w) + o(1), \quad w \neq 0, \tag{26}$$

for

$$T(B,w) = \prod_{j=1}^{d} \frac{\sin^2(\pi w_j \ell_j)}{(\pi w_j)^2}.$$

Proof. See Appendix B. □

This is not what we would expect as a large sample expectation of the spectrum, given our experience for time series and random fields in that the term $|B|^{-1}\lambda^2 T(B,w)$ has been added. To get there, we need to understand the DFT $J_h(w)$ better.

Proposition 4.1. *Assume that X is a stationary point process with intensity λ and with spectrum $f(w)$, and that h is a unit energy taper supported in the domain $B_\square$ itself only. Then the first moments of the direct spectral estimator $J_h(w)$ are given by:*

$$\begin{aligned}
\mathbf{E}\left\{J_h(w)\right\} &= \mathbf{E}\sum_{x\in X\cap B} h(x)\exp(-2\pi i w\cdot x) \qquad (27)\\
&= \lambda\int_{\mathbb{R}^d} h(x)\exp(-2\pi i w\cdot x)\,dx = \lambda H(w), \quad w\in\mathbb{R}^d\\
\mathbf{Var}\left\{J_h(w)\right\} &= \lambda + \lambda^2\int_{\mathbb{R}^d}|H(k'-w)|^2\,\widetilde{f}(k')\,dk' + \lambda^2|H(w)|^2 - \lambda^2|H(w)|^2\\
&= \lambda + \lambda^2\int_{\mathbb{R}^d}|H(k'-w)|^2\,\widetilde{f}(k')\,dk' \qquad (28)\\
\mathrm{Rel}\left\{J_h(w)\right\} &= \lambda\int_{\mathbb{R}^d} H(k'-2w)H(k')dk' + \int_{\mathbb{R}^d} U(w,z)e^{-2\pi i w\cdot z}\{\rho(z)-\lambda^2\}dxdz, \qquad (29)
\end{aligned}$$

with $U(w,z) = \int h(x)h(z+x)e^{-2\pi i w\cdot(2x)}\,dx$.

Proof. See Appendix C and notice the definition of (21). □

In the above Proposition the term Rel $\{\cdot\}$ denotes "relation", also known as the "complimentary covariance", see e.g. (Miller, 1974). These moments will have implications for the distributional properties of $I_h(w) = |J_h(w)|^2$. As we consider $w = 0$ then the variance and the relation are the same. For w small if $J_h(w)$ is Gaussian but the mean is non-zero and the moments are different than from the moments of a zero-mean time series and likely to lead to a non-centred χ^2 distribution.

As the DFT is the sufficient statistic (asymptotically) of a stationary process, in a sense, determining its first two moments are pivotal. Perhaps most important is (27) because of the non-central nature of the distribution. Understanding such distrubutions, e.g. the non-central χ^2 of the square of a Rician distribution, is not uncommon in medical imaging (Olhede and Whitcher, 2011). Inference is often not facilitated subsequently to making this observation–instead we shall propose to remove the mean, removing the non-centrality parameter.

We thus define a bias-corrected periodogram (see Daley and Vere-Jones 2003, p. 292, where the signed measure is used to make the equivalent definition), or

$$\widetilde{I}_h(w) = |J_h(w) - \lambda H(w)|^2 = \left|\widetilde{J}_h(w)\right|^2. \qquad (30)$$

Thus $J_h(w) - \lambda H(w)$ is often called the "signed measure". This is the DFT of the process X^0 dual to the mean-corrected random measure $N^0(dx) = N(dx) - \lambda dx$ where N is the dual of X. As we have subtracted $H(w)$ off the discrete Fourier transform, it is no longer strictly positive, but once we take the modulus square we are guaranteed to arrive at a real-valued positive quantity.

There is one wavenumber which is problematic, namely $w = 0$. For the periodogram we have $h_0(x) = \mathbf{1}_B(x)/\sqrt{|B|}$. Thus we have

$$J_0(w) = \frac{1}{\sqrt{|B|}}\sum_{x\in X\cap B} e^{-2\pi i w\cdot x}, \quad J_0(0) = \frac{1}{\sqrt{|B|}}N(B) = \sqrt{\|B\|}\widehat{\lambda}. \qquad (31)$$

This removes the problem of a non-zero mean of the DFT, as long as we assume that we know the intensity λ, and so are in the position to remove this effect. Assuming knowledge of this quantity is not a major hurdle, as it can be estimated consistently for largish areas ($|B|\to\infty$), by just counting the number of points and dividing by the area. For completeness we also define the signed measure DFT as

$$\widetilde{J}_h(w) = J_h(w) - \hat{\lambda}H(w), \quad . \qquad (32)$$

We note directly from (14), that

$$\widetilde{J}_h(w) \to N_C\{0, f(w), r(w)\}.$$

Also it follows

$$\widetilde{J}_0(0) = \sqrt{|B|}\widehat{\lambda} - \widehat{\lambda}H_0(0) = \sqrt{|B|}\widehat{\lambda} - \sqrt{|B|}\widehat{\lambda} = 0, \qquad (33)$$

trivially. Thus we cannot estimate the DFT at wave number zero. For a time series analysis when calculating DFTs we subtract off the sample mean, and then also the mean corrected DFT is zero at wave number zero. In a time series setting the periodogram at frequency zero is often not plotted.

We note that two ways of bias subtraction for the periodogram is possible, namely $\widetilde{I}(w)$ as described by (30) and additionally we could have defined

$$\breve{I}(w;h) = \sum_{x\in X\cap B}\sum_{y\in X\cap B} h(x)h(y)\exp\left\{-2\pi i w\cdot(x-y)\right\} - \widehat{\lambda^2}\left|H(w)\right|^2. \tag{34}$$

The advantage of the first estimator is that it is naturally non-negative and removes the error before we take the modulus square. This also should have a positive impact on the variance. In addition, $\widehat{\lambda^2}$ can be a biased estimator for λ^2.

If we are looking at a more sophisticated estimator than the separable estimator of (34) or (30) see e.g. (Cohen, 1995, Ch. 6), then we have

$$I_g(w) = \sum_{x\in X\cap B}\sum_{y\in X\cap B} g(x,y)e^{-2\pi i w\cdot(x-y)}, \tag{35}$$

where $g(x,y)$ is a *kernel* which may have a specified support, that has to be selected. For example we could define $g(x,y) = h(x)h^*(y)$ with $h(x)$ a multi-dimensional data taper (Hanssen, 1997), or we could make a non-separable choice of $g(x,y) = h(\|x-y\|)$. Depending on the choice of $g(x,y)$ the quantity $I_g(w)$ may satisfy a number of desirable characteristics. For example positivity, asymptotic unbiasedness as an estimator, and computational efficiency.

The quantity $\widetilde{I}_h(w)$ will be referred to as the "debiased periodogram", (and the subscript unity will be removed) and is preferable to use to the normal periodogram as it removes a number of unpalatable effects.

Lemma 4.2. *Assume that X is a homogeneous point process with a spectral density $f(w)$. Then the bias-corrected tapered periodogram has a first moment given by:*

$$\begin{aligned}\mathbf{E}\left\{\widetilde{I}_h(w)\right\} &= \mathbf{E}I_h(w) - 2\lambda\Re\left\{J_h(w)H^*(w)\right\} + \lambda^2\left|H(w)\right|^2\\ &= \int_{\mathbb{R}^d}\left|H(w'-w)\right|^2 f(w')dw'.\end{aligned} \tag{36}$$

Proof. See Appendix D. □

Thus as along as $H(w)$ is getting more concentrated in wave number (e.g. $H(w)\to\delta(w)$), this is asymptotically in $|B|$ an unbiased estimator of $f(w)$. Using the signed measure DFT of (32) to study X is more convenient, as this lets us avoid the contribution that affects the low wave numbers.

We now look to determine further properties of $I_h(w)$, or properties of $\widetilde{I}_h(w)$. To produce an estimator that is smoothed we need to study the covariance and variance of $\widetilde{I}_h(w_1)$ and $\widetilde{I}_h(w_2)$. We first look at $\mathbf{E}\left\{\widetilde{I}_h(w)\right\}$. We find that its form for large spatial regions is specified by the following theorem.

Let us write down what grid we get large sample unbiasesness for, customarily called the *Fourier grid*, and this is useful when the observational domain B is a box.

Definition 4.1. *The Fourier wave number grid for a point process observed on a cuboid domain $B_\square(\boldsymbol{l})$ corresponds to the points*

$$\mathcal{K}(\boldsymbol{l}) := \mathcal{K}(B_\square(\boldsymbol{l})) = \{w_n,\ w_n = \left(p_{n1}/l_1,\ldots,p_{nd}/l_d\right),\quad p_{nj}\in\mathbb{Z}\}. \tag{37}$$

Note that the physical units of the wave number elements is per unit length, such as m^{-1}. Referring to Lemma 4.1 we see that using this grid both removes any asymptotic bias at low wave numbers, but also gives us a sampling of the wave numbers that approximately leads to independent periodogram ordinates, rather like in the random field case. Note that the centering of the box $B_\square$ at zero, e.g. requiring the box to take the form

$$B_\square = \left(-\tfrac{\ell_1}{2}\quad\tfrac{\ell_1}{2}\right)\times\cdots\times\left(-\tfrac{\ell_d}{2}\quad\tfrac{\ell_d}{2}\right),$$

is required for the stated result. If the observational domain is not centred, then additional phase factors enter the expression. We can compare these results to wave number grid we chose to evaluate a standard spectral estimator of a time series, the Fourier frequencies (Percival and Walden, 1993b, p. 197–198). Their basic importance follows because we can expect the direct Fourier transform to be Gaussian and so uncorrelated implies independence.

A second feature of time series is the notion of the Nyquist wave number (Percival and Walden, 1993b, p. 98). This does not exist for point processes. It may seem counter intuitive that there is no upper limit to the

wave numbers we can estimate. When analysing a process that has been sampled in space, such as a random field, we expect to see aliasing. Aliasing is when variation that is happening very rapidly is confused with slower cycles, because when we have sampled more sparsely, rapid variation cannot be resolved. Because the distances can be any real-valued value, the Nyquist wave number does not exist.

Theorem 4.1. *Large–Domain Expectation of the tapered periodogram. Assume that X is a stationary point process with intensity λ and twice differentiable spectrum $f(w)$ at all values of $w \neq 0$, and that h is a unit energy taper (e.g. $\|h\| = 1$) supported in the cuboid domain $B_\square$ only. Assume X is observed in $B_\square$, which is growing in every dimension, that is $\min \ell_j \to \infty$. Then the expected value of the tapered periodogram $I_h(w)$ satisfies the equation*

$$\mathbf{E}I_h(w) = f(w) + \lambda^2 |H(w)|^2 + o(1), \quad w \neq 0, \tag{38}$$

where $H(w)$ is the Fourier transform of $h(x)$, as defined by Eqn. (21).

Proof. See Appendix E. □

In fact, asymptotics are well–understood both for time series and random fields. Our understanding of asymptotics for spectral estimation of point processes falls back to Daley and Vere-Jones (2003) or Diggle et al. (1987). In the latter reference Diggle et al. (1987) refers back to Chemistry theory by Hansen and McDonald, discussing the properties of gasses.

Corollary 4.1. *Assume that X is a stationary point process with intensity λ and twice differentiable spectrum $f(w)$ at all values of $w \neq 0$, and that h is a unit energy taper supported in the domain B only. Assume B is growing in every dimension. Then the expected value of the tapered periodogram $I_h(w)$ satisfies the equation*

$$\mathbf{E}I_h(w) = f(w) + \lambda^2 |H(w)|^2 + o(1), \quad w \neq 0, \tag{39}$$

where $H(w)$ is the Fourier transform of $h(x)$.

Proof. The proof can be found in Appendix G. □

This result is not unexpected and generalizes Theorem 4.1. The main difference here is that we consider a general domain B rather than a cuboid domain $B_\square$. This theorem allows us to determine the expectation of $\widetilde{I}_h(w)$, in general, as is made clear in the following corollary.

Corollary 4.2. *Bias–corrected periodogram. Assume that X is a stationary point process with intensity λ and twice differentiable spectrum $f(w)$ at all values of $w \neq 0$, and that h is a unit energy taper supported in the cuboid domain $B_\square$ only. Assume X is observed in $B_\square$, which is growing in every dimension. Then the expected value of the tapered periodogram $\widetilde{I}_h(w)$ satisfies the equation*

$$\mathbf{E}\widetilde{I}_h(w) = f(w) + o(1), \quad w \neq 0. \tag{40}$$

Proof. See Appendix F. □

This corollary shows the importance of removing the spectral bias before squaring–otherwise it gets hard to isolate and remove the effect at zero wave number–which is exacerbated at higher intensities (growing λ), as is clear from (38). So the estimator $\widetilde{I}_h(w)$ is debiased relative to $\mathbf{E}\{I_h(w)\}$ (as we shall see from Theorem 4.1), but is still guaranteed to be non-negative. This can be compared to removing $\widehat{\mu}$ from a time series before calculating the periodogram.

This gives us insight into the act of debiasing. In a sense our understanding of the spectral nature of point processes is informed by the spectral representation of randomly sampled processes (Chou and Lii, 1992). We can continue to explore the differences and similarities between spatial point processes and random fields.

Finally, to enable smoothing, define $P \geq 1$ estimates of the spectrum via

$$I_p^t(w) = \sum_{x,y \in X \cap B} h_p(x) h_p^*(y) e^{-i2\pi w \cdot (x-y)}, \quad p = 1, \dots P, \tag{41}$$

where we assume

$$\int_{\mathbb{R}^d} h_p(x) h_l^*(x) dx = \delta_{pl}$$

where $\delta_{pl} = 1$ if $p = l$ and 0 otherwise. These estimates can also be arrived at starting from (13) and $J_h(w)$. As we shall use P tapers linearly, and P^2 for the quadratic estimators, we shall use the subscript '0' to refer to the untapered periodogram, and $p = 1, \dots, P$ to refer to the subsequent tapers.

Tapering is very common for stochastic processes. Initially the idea was hotly contested, but its utility is now firmly established both for time series and random fields (Dahlhaus and Künsch, 1987; Walden, 2000). The idea

of tapering is to ameliorate edge effects that leads to the asymptotically leading bias term. This is important in 1-D (Thomson, 1982; Walden, 2000) but of greater importance in 2-D and higher (Dahlhaus and Künsch, 1987). In 2-D and higher the edge effects become more pronounced, and indeed asymptotically dominant (Kent and Mardia, 1996; Mardia and Marshall, 1984). This paper also uses multitapering to stabilize the variance of any estimator, as described for time series in (Walden, 2000). We shall describe the necessary steps to perform linear estimation of the spectrum of a 2-D and higher dimension point process.

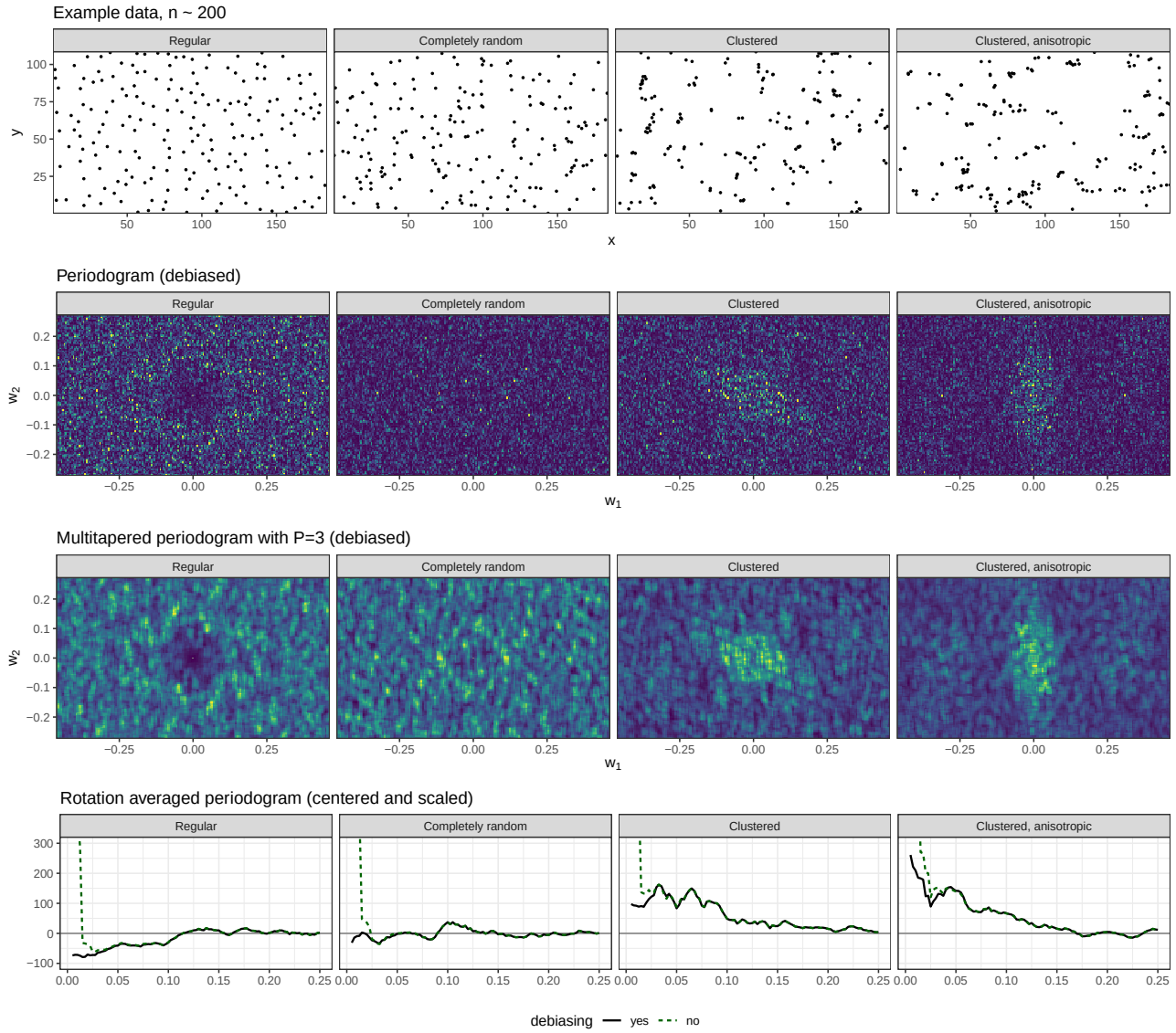

Figure 2: *Top row*: Four example data patterns in a rectangle of area 2e4 units, exhibiting regularity, complete randomness, small scale clustering, and clustering with a left-right axis directionality. *Second row*: Their debiased periodograms excluding $w = 0$. *Third row*: Their multitapered debiased periodograms, 3×3 sine-tapers. *Bottom row*: Their rotationally averaged periodograms, including the case without debiasing. The 2D summaries use a w-grid with a step size 0.005 in both directions. The rotational averaging was estimated at $|w| = 0.0050, 0.0075, ..., 0.2500$ using a box kernel having a radius 1.5 times the w-grid step size 0.005.

Let us set theory aside for a moment and have a look at some spectral estimates: Figure 2 demonstrates the debiased periodogram, the debiased multitapered periodogram with sine-tapers and a rotationally averaged 1D summary of the peridogram (Sections 6&7) for four point patterns exhibiting different structural behaviour. The aspect ratio of the plots are kept the same only to have an orderly figure: Due to debiasing we can estimate the periodogram on a wavenumber grid of our own choosing, and not just on the Fourier grid (which in these finite examples would be different for different tapers, cf. Section 5). The wavenumber-grid in this illustration is a regular grid with a step size 0.005 in both dimensions.

We see that information is present at wavenumbers up to $|w| \approx 0.2$. The tapering smooths the periodogram, as expected. There is a dark well near the origin for the regular pattern, and a bright hump for the clustered pattern, and both features transfer to the rotationally averaged curves (compare with Figure 1). The anisotropy

of the fourth pattern is hinted by the anisotropy visible in the 2D periodograms as elongation of the "hump" in the second dimension. The elongation is perpendicular to the elongations of the clusters in the data because wavenumbers relate inversely to spatial units. The rotationally averaged periodogram was also computed from the (not shown) non-debiased periodogram for comparison. There is a very prominent bias near $|w| = 0$ for the non-debiased version, which hopefully illustrates why the proposed debiasing step is relevant when applying the method in practice on freely chosen wavenumber grids.

5 Variance and Covariance of Bilinear Spectral Estimators

Before determining how to do bilinear estimation of the spectral density let us determine the second order properties of spectral estimators. Core to our understanding of smoothing will be the variance and covariance of the tapered functions. This will be established in the following theorem. For brevity define

$$\phi_w(x) = \exp\{-2\pi i w \cdot x\}, \quad w, x \in \mathbb{R}^d. \tag{42}$$

With this notation we have

$$I_p^t(w) = \sum_{x,y \in X \cap B} h_p(x) h_p^*(y) \phi_w(x-y), \quad p = 1, \dots P. \tag{43}$$

Above we note that we have the same value of p across x and y in (43) as this corresponds to a modulus square.

Theorem 5.1. *Let $I_p^t(k)$ denote the (tapered) periodogram given by (41). The covariance between the (tapered) periodogram and itself across two wavenumbers $w_1, w_2 \in \mathbb{R}^d$ then takes the form of*

$$\begin{aligned}
\mathbf{Cov}\{I_p^t(w_1), I_l^t(w_2)\} &= \int_{B^4} \rho^{(4)}(x,y,z,v) h_p(x) h_p(y) h_l(z) h_l(v) \phi_{w_1}(x-y) \phi_{w_2}(z-v)\, dxdydvdz \\
&+ \int_{B^3} \rho^{(3)}(x,y,v) \left[h_p(x) h_p(y) h_l(x) h_l(v) \phi_{w_1}(x-y) \left\{ \phi_{w_2}(x-v) + \phi_{-w_2}(x-v) \right\} \right. \\
&+ h_p(x) h_p(y) h_l(y) h_l(v) \phi_{w_1}(x-y) \left\{ \phi_{w_2}(y-v) + \phi_{-w_2}(y-v) \right\} \\
&\left. + \phi_{w_2}(x-y) h_p^2(v) h_l(x) h_l(y) + \phi_{w_1}(x-y) h_l^2(v) h_p(x) h_p(y) \right]\, dxdydv \\
&+ \int_{B^2} \rho^{(2)}(x,y) \left[h_p(x) h_p(y) h_l(x) h_l(y) \phi_{w_1}(x-y) \left\{ \phi_{w_2}(x-y) + \phi_{-w_2}(x-y) \right\} \right. \\
&\left. + \left\{ h_p^2(x) h_l(x) h_l(y) + h_l^2(x) h_p(x) h_p(y) \right\} \left\{ \phi_{w_1}(x-y) + \phi_{w_2}(x-y) \right\} + h_p^2(x) h_l^2(y) \right]\, dxdy \\
&+ \lambda \int_B h_p^2(x) h_l^2(x)\, dx - \lambda^2 - \lambda \int_{B^2} h_p(x) h_p(y) h_l(x) h_l(y) (\phi_{w_1}(x-y) + \phi_{w_2}(x-y)) \rho^{(2)}(x,y) dxdy \\
&- \left\{ \int_{B^2} h_p(x) h_p(y) \phi_{w_1}(x-y) \rho^{(2)}(x,y) dxdy \right\} \cdot \left\{ \int_{B^2} h_l(x) h_l(y) \phi_{w_2}(x-y) \rho^{(2)}(x,y) dxdy \right\}.
\end{aligned}$$

Proof. The proof is in Appendix H. □

Note that we can in fact write:

$$\widetilde{I}_h(w) = I_h(w) - 2\lambda \Re\left\{ J_h(w) H^*(w) \right\} + \lambda^2 \left| H(w) \right|^2.$$

The expression derived in Theorem 5.1 is quite general, and only requires the assumptions of homogeneity in space, but it is hard to understand, especially as it is hard to order the terms in size without putting very restrictive assumptions in place. We can still consider the untapered case by setting $p = l$ and taking $h(x) = |B|^{-1/2} \mathbf{1}_B(x)$, to arrive at an expression. Let us study this general correlation in the instance of the Poisson process where we know that

$$\rho^{(n)}(x_1, \dots, x_n) = \lambda^n, \tag{44}$$

and consider the case of $w_1 = w_2 = w$.

For completeness we also define the spectral bandwidth b_h by

$$b_h^2 = \int_{\mathbb{R}^d} \|w\|^2 \left| H(w) \right|^2 dw. \tag{45}$$

Proposition 5.1. *Let $I_p^t(w)$ denote the tapered periodogram given by (41) for a Poisson process with intensity λ. The covariance between the tapered periodogram with one taper and the tapered periodogram with another taper then takes the form for $\|k\| > b_h$ of*

$$\mathbf{Cov}\{I_p^t(w), I_l^t(w)\} = \lambda \int_B h_p^2(x) h_l^2(x)\, dx + \lambda^2 \delta_{pl} + o(1).$$

The form for $\|w\| < b_h$ is much more complex, and is provided in Appendix J.

Proof. The proof can be found in Appendix J. □

Remark 5.1. *At this stage we have simplified our assumptions to the Poisson process which is quite disappointing. The reason why we have chosen to do so is clear from Theorem 5.1. If we assume the process is Poisson then a lot of terms cancel. The correct treatment of non-Poisson processes would correspond to transforming the terms into a wavenumber representation, doing a Taylor series because of the concentration of the tapers, and then implementing the same calculation after the Taylor series.*

Proposition 5.2. *Let $I_p^t(w)$ denote the tapered periodogram given by (41) for a Poisson process with intensity λ. The covariance between the (tapered) periodogram and itself at different wave numbers takes the form of*

$$\mathbf{Cov}\{I_p^t(w_1), I_p^t(w_2)\} = \lambda \|h_p\|_4^4 = o(1).$$

Proof. The proof can be found in Appendix K. □

Fix $0 < \epsilon < 1$. This is so that we can approximately find an uncorrelated grid. Recall that $h(x)$ is unit energy in space or wave number.

Definition 5.1. *The tapered Fourier wave number grid for a point process observed on a cuboid domain $B_{\square}(\boldsymbol{l})$ corresponds to the points*

$$\mathcal{K}(B_{\square}(\boldsymbol{l})) = \{w_n,\ w_n = \left(p_{n1}\tau(\epsilon)/l_1, \ldots, p_{nd}\tau(\epsilon)/l_d\right), \quad p_{nj} \in \mathbb{Z}\}, \tag{46}$$

where τ is defined from the equation

$$\min_{\tau(\epsilon)} \left| \int H_p(\tau(\epsilon) - w') H_l^*(w')\, dw' \right| < \epsilon.$$

These results hint at producing general methodology; we can for the Poisson process figure out how to do non-parametric estimation. This is however not enough. We need to determine the variance of the periodogram in more general. The following proposition establishes its form.

Theorem 5.2. *Assume X is a stationary process. The variance of the (tapered) periodogram then takes the form of*

$$\mathbf{Var}\{I_p^t(w)\} = \lambda^4 \left\{ \widetilde{f}^{(4)}(w,w,w) - \widetilde{f}^{(2)}(w)(1 + 2\widetilde{f}^{(2)}(w) \right\} + \lambda^2 \widetilde{f}(w) + \lambda^2 + \lambda \|h\|_4^4 + o(1).$$

Proof. The proof can be found in Appendix L. □

With these moments we can now consider the problem of estimation. We might feel an increasing degree of unease as we have succumbed to the usual fallacy of only deriving the most general results for a Poisson process. However as the observational domain becomes larger we would expect $|H_p(w)|$ to concentrate and so the lack of constancy will become less of an issue.

6 Linear Smoothing

Having derived the first and second order properties of the periodogram, the next task is then estimation of $f(w)$, as the primal quantity of interest in studying the point process X. If we fully characterise $f(w)$ then we have fully characterised $\gamma(z)$, as the Fourier transform is a bijection. Just like for a time series or random field, a key question concerns the correct method for quadratic characterisation and estimation.

For the process itself (X), then its first moment is encapsulated by λ for a stationary (or homogeneous) processes. We recall the spectrum uniquely maps to the complete covariance function $\gamma(z)$. We want to start from the random variables $\{I_h(w)\}$ (or $\{\widetilde{I}_h(w)\}$) and then non-parametrically estimate the spectral density directly from those quantities. As long as we seek to estimate $f(w)$ at points w where the function $f(w)$ is continuous, averaging $\{I_0(w)\}$ (or $\{\widetilde{I}_h(w)\}$) locally seems like a sensible strategy. (Diggle et al., 1987) proposed smoothing $\widetilde{I}_D(w)$ as defined in (19). They just note that they apply a weighted average smoother with user tuned weights. This will not determine how to sample wavenumbers, or remove correlation, and we also have the alternative form of bias removal to them before smoothing.

We shall only treat growing domain scenarios ($|B| \to \infty$ (one can contrast growing domain observations and increasing density of observations), see e.g. Zhang and Zimmerman (2005) or Stein (1995). Of course when faced with a real point pattern we cannot tell if asymptotic methods offer a good approximation to the finite sample distribution.

For sufficiently high wave numbers the constant effect in the spectral density with expectation λ will dominate both in the mean and variance of $I_0(w)$. Thus for higher wave numbers in w the periodogram will eventually

become perfectly correlated over wavenumbers, and will be converging to $\widehat{\lambda}$. For a growing domain we can show that $\hat{\lambda}$ converges in probability to λ.

Our understanding of the correlation of higher wave-numbers should be contrasted with the discussion of Mugglestone and Renshaw (1996a). The latter authors remark that there is an upper limit on the number of independent periodogram ordinates, see either of Mugglestone and Renshaw (2001, 1996b), but do not justify that statement (which is based on random field analysis rather than point processes). It is not clear from these manuscripts exactly how this upper limit is arrived at, as Mugglestone and Renshaw (2001) it is based on heuristic (or one could say practical) arguments on the nominal degrees of freedom present in the data, in the random field limit.

We have already described how to de–bias $J_h(w)$ into $\widetilde{J}_h(w)$, which is a properly zero–mean quantity. We clearly do not need to do this as long as we stick to the Fourier grid. Mean correction has also been discussed by (Daley and Vere-Jones, 2003, p.332) to achieve analogous results to the spectral representation of time series. This mean correction has the advantage of letting us determine the properties of $J_h(w)$ directly, and as we chose to evaluate Bartlett's rather than Diggle's spectral estimator, we have the complex–valued quantity available to us, and can do the mean correction while still retaining a real-valued positive spectral estimate. Mean correction for the periodogram after squaring is more precarious, as we are not guaranteed a positive estimate.

Having understood the periodogram point-wise in terms of its expectation, to understand how to smooth it, we also need to make sense of its variance/covariance structure. Smoothing requires an explicit understanding of local degrees of freedom. For this purpose we shall use tapered estimates, see also (Walden, 2000). We refer to the previous section and (41) for the basic taper definitions.

We have two possible estimators, namely the bias–corrected multi-taper estimator of (Walden, 2000) as well as a multi-dimensional kernel density estimator (Duong and Hazelton, 2005). They are both linear in the estimated spectrum. We shall use the periodogram rather than another estimator for the kernel density estimation, as it is easier to keep track of bias and correlation.

We therefore define the multitaper estimator (Walden, 2000) (and refer to (Hanssen, 1997; Liu and Van Veen, 1992))

$$\overline{I}_P^t(w) = \frac{1}{P}\sum_{p=1}^{P} \widetilde{I}_{h_p}(w), \quad w \in \mathbb{R}^d, \tag{47}$$

where $\widetilde{I}_{h_p}(w)$ is defined in (30). Choosing to use continuous tapers in (30) is not the only option, and this definition allows for nonseparable tapers (Simons and Wang, 2011), as it does not loop over Cartesian product tapers. We could have calculated tapers for our spatial samples, see (Wingham, 1992; Bronez, 1988), or we could have interpolated the points (Das and Babadi, 2020). Neither solution is palatable, and so we have stuck with continuous space tapers evaluated at random locations. This is a deliberate choice, as otherwise our method would become opaque to analysis.

If we use tapers then to avoid bias, it is necessary to debias $\widetilde{I}_{h_p}(w)$. If we stick to the Fourier grid, then the bias is zero and not necessary to remove. Thus starting from multi-dimensional kernel density estimators (Duong and Hazelton, 2005) we instead chose a window function $W(x)$ and matrix bandwith $\mathbf{\Omega}$. We then define

$$\overline{I}_h(w;\mathbf{\Omega}) = \sum_{w'\in\mathcal{K}} |\mathbf{\Omega}|^{-1} W\left(\mathbf{\Omega}^{-1}\left(w'-w\right)\right) \widetilde{I}_h\left(w'\right), \tag{48}$$

where $\widetilde{I}_h\left(w\right)$ is defined by (30). The 'h' just denotes the taper, and the tilde captures that we are using a bias-corrected periodogram. A useful modification is to consider one-dimensional window W and $\mathbf{\Omega} = \Omega$ and consider the averaging over directions, thus reducing the statistic to magnitude $t = \|w\|$ only,

$$\overline{I}_h(t;\Omega) = \sum_{w'\in\mathcal{K}} \Omega^{-1} W\left(\Omega^{-1}\left(||w'|| - t\right)\right) \widetilde{I}_h\left(w'\right), \tag{49}$$

which we refer to as the rotationally averaged periodogram. Mugglestone and Renshaw (1996b) call a version of it the R-spectrum. Alternatively, one could average also over the magnitudes to arrive at a summary function in direction. In 2D the convenient argument would be the polar angle; Mugglestone and Renshaw (1996b) call this version the θ-spectrum.

Basically both the multi-tapering and the local averaging smooth any direct spectral estimate (or bilinear estimator). Both of them have a smoothing parameter, namely the number of tapers P or the smoothing matrix bandwidth $\mathbf{\Omega}$. Let us determine how we select them, and discuss how this might be done in theory as well as practice. We shall focus on 2D as this is the best explored dimensionality after 1D.

6.1 Bandwidth selection in 2D

We have already discussed the tapering in (43). Some care has to be taken over the indices we loop over, especially the spatial index and the taper index. Eqn (43) has to contain the same index over the taper as it

is a modulus square. On the other hand in (50) we loop over the taper indices to combine the 1-d Cartesian tapers to make 2-d Cartesian tapers, e.g. we would use 00 then 01, 10 as well as 11. The two types of indices should not be confused.

With this, we are now ready to turn our theoretical understanding of $I_p^t(k)$, the tapered periodogram, into practical methodology. Starting from the multiple estimates of the spectrum we define the multitaper estimates for $x \in B \in \mathbb{R}^2$

$$\overline{I}^t_{P_1P_2}(w) = \frac{1}{P_1P_2}\sum_{p_1=1}^{P_1}\sum_{p_2=1}^{P_2}\widetilde{I}_{h_{p_1p_2}}(w), \quad w \in \mathbb{R}^2. \tag{50}$$

Note that when looking at 1D tapers p ranges from 1 to P_1 but when going higher dimension we need to loop over each dimension in turn. The higher the dimension, the more complex are our options. In 2D we shall simply explore Cartesian products of 1D tapers.

The number of tapers P_1 and P_2 in each of the two dimensions in (50) needs to be selected to balance variance and bias. For intuition we shall first understand the mean square error of a local averaging procedure as that defined in (48). We shall thus be inspired by regular d-dimensional smoothing to determine how to smooth. Starting from standard understanding available for d-dimensional density estimation (Duong and Hazelton, 2005; Wand and Jones, 1993) we use the kernel function $W(k)$ with matrix-bandwidth $\mathbf{\Omega}$. Scaling the kernel function using the bandwidth matrix we obtain

$$W_{\mathbf{\Omega}}(w) = |\mathbf{\Omega}|^{-1} \cdot W\left(\mathbf{\Omega}^{-1}w\right), \quad w \in \mathbb{R}^2. \tag{51}$$

The bandwidth function satisfies

$$\sum_{w \in \mathcal{K}(B_\square(\boldsymbol{\ell}))} W(w) = 1. \tag{52}$$

Note that this is not the same as the continuous version of the taper integrating to unity, see also the Section discusses the tapers.

With this definition we get the kernel density estimator

$$\overline{I}_h(w;\mathbf{H}) = \sum_{w' \in \mathcal{K}(B_\square(\boldsymbol{\ell}))} \widetilde{I}_h(w')W_{\mathbf{\Omega}}(w-w'), \quad w \in \mathbb{R}^d. \tag{53}$$

Note that (50) is not a special case of (53). As each quadratic estimate $I_p^t(w)$ is arrived at by linear operations on the DFT, and then implementing a modulus square, not all direct spectral estimates $I_p^t(w)$ can be arrived at from smoothing the periodogram $I_0(w)$.

Lemma 6.1. *Assume $\widetilde{f}(w)$ for all $w \neq 0$ is twice differentiable in w, and denote the mixed second derivative by $\frac{\partial^2 \widetilde{f}}{\partial w_i \partial w_j} = \widetilde{f}_{ij}(w)$. Then the absolute bias of (53) as an estimator of $f(w)$ is given for $w \neq 0$ by*

$$\left|\text{Bias}_0\left\{\overline{I}_h(w;\mathbf{H})\right\}\right| \leq \frac{\lambda^2}{2}\max_i |\widetilde{f}_{ii}(w)|\tau_h^2, \tag{54}$$

$$\tau_h^2 = \sum_{w' \in \mathcal{K}(B_\square(\boldsymbol{\ell}))} (w') \cdot w' W_h(w').$$

Proof. See Appendix N. □

The next step is the variance and the mean square error of our smoothed spectral estimator.

Lemma 6.2. *Assume $\widetilde{f}(w)$ for all $w \neq 0$ is twice differentiable. Then the variance of (53) as an estimator of $f(w)$ is given for $w \neq 0$ by*

$$\mathbf{Var}\{\overline{I}_h(w;\mathbf{\Omega})\} = \frac{f(w_n) - \lambda + \lambda^2 + \lambda\|h\|_4^4}{\sigma_k \ell} + \lambda\|h\|_4^4. \tag{55}$$

Proof. See Appendix O. □

We put these results together and arrive at the optimal bandwidth.

Theorem 6.1. *We assume that X is a point process with spectrum $f(w)$ that is twice differentiable for all $w \neq 0$. Assume we smooth the periodogram using a box-car window with width σ, and take $\tilde{\sigma} = \sigma\ell h$ as the bandwidth measures in terms of the length ℓ. In this case the mean square error takes the form of*

$$\text{MSE}\left\{\widetilde{I}(w_n)\right\} \leq \frac{\lambda^4}{4}\max_i |\widetilde{f}_{ii}(w')|^2\left\{d(\tilde{\sigma}^2)\frac{1}{12}\right\}^2 + \frac{f(w_n) - \lambda + \lambda^2 + O(|B|^{-1})}{\tilde{\sigma}} + O(|B|^{-1}).$$

This mean square error is minimized at

$$\tilde{\sigma} = \left(12^2 \frac{\lambda^2}{d^2 \lambda^4 \max_i |\widetilde{f}_{ii}(w')|^2} \right)^{1/5}. \tag{56}$$

Proof. See Appendix P. □

Naturally for a real data set, the parameters in this expression must be estimated, and the second derivative depends on the quantity we want to estimate. Given the theorem shows that an optimal smoothing bandwidth exists, it makes sense to pick one, and we propose, as we shall demonstrate in simulation studies, that visually a suitable degree of smoothing can be arrived at in examples. Thus the main point of this section is that an optimal wavenumber can be found, and it depends on the intensity of the point process as well as the roughness of the spectrum.

6.2 Multitaper number

This is all very well to find the bandwidth for the locally averaged estimator specified by (53), but we also want it for the multi–taper estimator in (50). The two most popular sequences of tapers to use are the prolate spheroidal wave sequences (Landau and Pollak, 1962) and the sinusoidal functions (Riedel and Sidorenko, 1995). The former is used because it is maximally concentrated in wave number whilst being compact in time/space; the second because it produces the least bias formally.

Note that our choice is not trivial; some authors chose to interpolate the point process to a regularly spaced process, and then taper. We could chose to design taper to match the observed locations, e.g. Bronez (1988), but the analysis of that choice would be complicated.

We shall use P^d tapers in d dimensions, and in this case we can derive the same expression for the bias as in (54) but we now need to use the properties of the discrete prolate spheroidal eigensequences–the number of well-concentrated tapers with a bandwidth σ is $P = 2\ell\sigma$ for a cuboid data analysis window (Percival and Walden, 1993b, p 369). We then get the same bias expression for each tapered periodogram as for $k > \sigma$ of

$$\mathbf{E}\left\{ I_p^t(w;h) \right\} = \lambda + \lambda^2 \int_{\mathbb{R}^d} |H_p(w'-w)|^2 \, \widetilde{f}(w') \, dw', \quad p = 1, \ldots, P^d. \tag{57}$$

We do this by mapping the numbering of each dimension into a single enumeration. Having determined the expectation of each term we find that

$$\begin{aligned} \mathbf{E}\overline{I}_P(w) &= \frac{1}{P^d} \sum_{p=1}^{P^d} \mathbf{E} I_p^t(w) \\ &= \lambda + \lambda^2 \int_{\mathbb{R}^d} \frac{1}{P^d} \sum_{p=1}^{P^d} |H_p(w'-w)|^2 \, \widetilde{f}(w') \, dw', \end{aligned}$$

this defining the spectral window of the taper by

$$\mathcal{H}_P(w) = \frac{1}{P^d} \sum_{p=1}^{P^d} |H_p(w)|^2, \; p = 1, \ldots, P^d.$$

Using spheroidal wave functions, assuming the observational spatial domain is a Cartesian product of intervals of length ℓ, we find that (Percival and Walden, 1993b, p 342) with $P = 2\ell\sigma$ then

$$\frac{1}{P^d} \sum_{p=1}^{P^d} |H_p(w'-w)|^2 \approx \begin{cases} 1 & \text{if} \quad \forall \, |w_j| < \sigma \\ 0 & \text{if} \quad \exists \, j : \; |w_j| < \sigma \end{cases}. \tag{58}$$

This follows directly from the choice of direct product domains, and the use of the spheroidal wave functions.

We again expand the spectral density function to

$$\widetilde{f}(w') = \widetilde{f}(w) + \nabla \widetilde{f}(w) \cdot (w'-w) + \frac{1}{2}(w'-w) \cdot \widetilde{\boldsymbol{H}}(w'')(w'-w). \tag{59}$$

We may as before note that

$$\begin{aligned} \mathbf{E}\overline{I}_P(w) &= \lambda + \lambda^2 \int_{\mathbb{R}^d} \mathcal{H}_P(w'-w) \left[\widetilde{f}(w) + \nabla \widetilde{f}(w) \cdot (w'-w) + \frac{1}{2}(w'-w)^T \widetilde{\boldsymbol{H}}(w'')(w'-w) \right] dw' \\ &= \lambda + \lambda^2 \widetilde{f}(w) + \frac{\lambda^2}{2} \int_{\mathbb{R}^d} \mathcal{H}_P(w'-w)(w'-w)^T \widetilde{\boldsymbol{H}}(w'')(w'-w) \, dw'. \end{aligned}$$

We may argue about the bias in a manner analogous to before. We find:

$$\begin{aligned}|\text{bias}\left\{\overline{I}_P(w)\right\}| &\leq \frac{\lambda^2}{2}\int_{\mathbb{R}^d}\mathcal{H}_P(w'-w)\left|(w'-w)^T\widetilde{\boldsymbol{H}}(w'')(w'-w)\right|\,dw' \\ &\leq \frac{\lambda^2}{2}\max_{k,i}|\widetilde{f}_{ii}(w')|^2\int_{\mathbb{R}^d}\mathcal{H}_P(w'-w)\sum_{i=1}^{d}(w'-w)_{ii}^2\,dw' \\ &= \frac{\lambda^2}{2}\max_{w',i}|\widetilde{f}_{ii}(w')|^2\cdot d\cdot\frac{1}{12}\sigma^2.\end{aligned}$$

We now use that $P = 2\ell\sigma$. For this reason our bias remains the same as before:

$$|\text{bias}\left\{\overline{I}_P(w)\right\}| \leq \frac{\lambda^2}{2}\max_{w',i}|\widetilde{f}_{ii}(w')|^2\cdot d\cdot\frac{1}{12}(P/\{2\ell\})^2.$$

The variance of the estimator comes from individual contributions from covariance expressions:

$$\mathbf{Var}\{\overline{I}_P(w)\} = \frac{1}{P^d}\sum_{p=1}^{P^d}\frac{1}{P^d}\sum_{q=1}^{P^d}\mathbf{Cov}\{I_p^t(w), I_q^t(w)\}.$$

To approximate the covariance we observe Theorem 5.2 which gives the form of the variance, where we ignore the higher order terms.

We then note

$$\mathbf{Cov}\{I_p^t(w), I_q^t(w)\} \asymp \delta_{pq}\left\{\lambda^2(1+\tilde{f}(w)) + \lambda\mathcal{O}(\|h\|_4^4)\right\} + \lambda\mathcal{O}(\|h\|_4^4). \tag{60}$$

Because of the scaling we get

$$\int_B h_p^4(x)\,dx = \mathcal{O}\left(\frac{1}{|B|}\right), \tag{61}$$

and so putting these together we get that

$$\begin{aligned}\mathbf{Var}\{\overline{I}_P(w)\} &= \frac{1}{P^{2d}}\sum_{p=1}^{P^d}\sum_{q=1}^{P^d}\left\{\delta_{pq}\left\{\lambda^2(1+\tilde{f}(w)) + \lambda\mathcal{O}(\frac{1}{|B|})\right\} + \lambda\mathcal{O}(\frac{1}{|B|})\right\} \\ &= \frac{\lambda^2(1+\tilde{f}(w)) + \mathcal{O}(\frac{1}{|B|})}{P^d} + \mathcal{O}\left(\frac{1}{|B|}\right).\end{aligned}$$

We can then write

$$\begin{aligned}\text{MSE}\{\overline{I}_P(w)\} &= \left\{\frac{\lambda^2}{2}\max_{w',i}|\widetilde{f}_{ii}(w')|^2\cdot d\cdot\frac{1}{12}(P/\{2\ell\})^2\right\}^2 + \frac{1}{P^d}\left\{\lambda^2(1+\tilde{f}(w)) + \mathcal{O}\left(\frac{1}{|B|}\right)\right\} + \mathcal{O}\left(\frac{1}{|B|}\right) \\ &= \frac{\lambda^4}{4}\max_{i,w'}|\widetilde{f}_{ii}(w')|^2 d^2P^4/\{16\ell^4\} + \frac{1}{P^d}\left\{\lambda^2(1+\tilde{f}(w))\right\} + \mathcal{O}\left(\frac{1}{|B|}\right).\end{aligned}$$

We can yet again minimize in P and so use calculus to do so

$$\begin{aligned}\frac{\partial}{\partial P}\text{MSE}\{\overline{I}_P(w)\} &= \frac{\lambda^4}{16}d^2\max_{w',i}|\widetilde{f}_{ii}(w')|^2P^3/\ell^4 - \frac{d}{P^{d+1}}\left\{\lambda^2(1+\tilde{f}(w)) + \mathcal{O}\left(\frac{1}{|B|}\right)\right\} \\ \Rightarrow P^{d+4} &= \frac{16\left\{\lambda^2(1+\tilde{f}(w)) + \mathcal{O}\left(\frac{1}{|B|}\right)\right\}\ell^4}{\lambda^4 d^2\max_{w',i}|\widetilde{f}_{ii}(w')|^2}.\end{aligned} \tag{62}$$

Again we see how the optimal taper number scales. There is a clear problem with the "curse of dimensionality", as the range of P decays with d. This indeed gives a very similar answer to our previous smoothing, and so we can estimate how many taper we need to use using the previous parameters. Again, the choice of taper necessitates knowledge of the second derivative of what we are trying to estimate–if we truly knew the second derivative we would not need to implement that estimation. Instead what the proof here shows is the scaling in the fundamental parameters of the process, and that a balance between variance and bias can be made. Furthermore this is not trivially done at the boundary of using P^d or 1 taper.

In practice we do not have the unknowns in equations (56) and (62). But the advantage of determining their value is that we know that an optimum exists: we shall therefore empirically look for it using techniques applied in time series. An approach in time series is to pre-select the wave number resolution $2W$,

see for example (Percival and Walden, 1993a, p. 369). The authors note that if the basic resolution is W then we select $K = 2NW\Delta t - 1$. In fact for a time series sampled at Δt steps apart we call $2NW\Delta t$ the Shannon number (Percival and Walden, 1993a, p. 335). If $2W$ is the bandwidth in wave number of the data-taper concerntation operator, then $k = 0, ..., K-1$ where $K < 2NW\Delta t$ will yield a set of tapers $\{h_k(\cdot)\}$ that are optimally concentrated to the wave number band $[-W\Delta t, W\Delta t]$. So one method of selecting the number of data tapers is to just select the wave number resolution w, and then take the value of K that matches the Shannon number.

6.3 Plotting Spectral Estimates

For regularly spaced time series it is very clear what frequencies/wave numbers we plot. We stick to the Fourier wave numbers, and plot up to the Nyquist wave number. But for a point process no such upper limit exists, as is also the case for randomly sampled stochastic processes, see (Lii and Masry, 1994).

What we will find at larger wave numbers is that the baseline contribution of λ dominates the spectrum. Unlike time series that are regularly sampled in time, and it is natural to assume that the spectrum of the process is decaying in wave number, we can only assume the spectral deviations $\widetilde{f}(w)$ are decaying with $\|w\|$.

How do we then determine what is the highest wave number we can do spectral estimation? To determine this number we must guess at the decay of $\widetilde{f}(w)$. We shall assume that

$$\widetilde{f}(w) = C_0\|w\|^{-\alpha}, \quad \|w\| > \|w'\| \tag{63}$$

If we knew $\|w'\|$ then our approach would be simple. We would take

$$\begin{aligned} I(w_k) &= \hat{\lambda} + \hat{\lambda}^2 C_0\|w_k\|^{-\alpha}, \quad k = 1, \ldots, K \\ Y_k &= I(w_k) - \hat{\lambda} = \hat{\lambda}^2 C_0\|w_k\|^{-\alpha}, \quad k = 1, \ldots, K \\ Z_k &= \log|I(w_k) - \hat{\lambda}| = 2\log\hat{\lambda} + \log C_0 - \alpha\log\|w_k\|. \end{aligned} \tag{64}$$

We can then assign $\|w_1\|$ as the wave number taking the value $\frac{1}{2}I(\frac{1}{\ell_1}, \frac{1}{\ell_2})$ up to that taking the value $2\widehat{\lambda}$ using the Fourier grid and then solve for (C_0, α) using least squares, this giving us $(\hat{C}_0, \hat{\alpha})$. With these estimates we solve

$$\hat{C}_0\|w_h\|^{-\hat{\alpha}} = \widehat{\lambda},$$

and use this as the highest wave number, w_h.

7 Isotropy versus Anisotropy

We now consider the special case of isotropic point processes for which the complete covariance function depends only on distance, $\gamma(z) = \gamma_I(||z||)$ (see also Daley and Vere-Jones, 2003, p. 310–311). Such radial functions (Grafakos and Teschl, 2013) are special as we do not require any orientational specificity in our spectral representation. The isotropy of $\gamma(z)$ transfers through the Fourier transform to the sdf so $f(w) = f_I(||w||)$. We shall refer to the treatment of isotropic random fields when implementing analysis (Ponomarenko and Perun, 2006). An alternative is discussed by (Errico, 1985) and (Durran et al., 2017), for random fields.

The orientation invariance leads to dimensional reduction of the multidimensional Fourier transform. Three basic concepts are useful for the isotropic analysis: Recall the Bessel function of order $m > -1/2$

$$\mathcal{J}_m(t) = \frac{t^m}{2^m\pi^{1/2}\Gamma(m+1/2)}\int_0^\pi \sin^{2m}(u)\ \cos[t\ \cos(u)]\ du \qquad t \geq 0.$$

It is connected to the d-dimensional Fourier transform via (Vembu, 1961, Sec 3)

$$\int_{S^{d-1}} e^{-2\pi l i v\cdot u} du = 2\pi\ l^{-(d/2-1)}\mathcal{J}_{d/2-1}(2\pi l) \quad \forall v \in S^{d-1}, l \geq 0,$$

we can also note relationships between radial and Cartesian Fourier representations, see for example Grafakos and Teschl (2013), or Estrada (2014).

Recall also the (1-dimensional) Hankel transform of order m of a function $g : [0, \infty] \mapsto \mathbb{R}$

$$\mathcal{H}_m[g](l) = \int_0^\infty g(r)\mathcal{J}_m(lr)r\ dr \qquad l \geq 0.$$

Then if we assume X is a stationary and isotropic process with a radial complete covariance function γ_I, the sdf at wavenumber w with $t = \|w\|$ takes the form

$$\begin{aligned} f_I(t = \|w\|) := f(w) &= \lambda + \int_{\mathbb{R}^d} \left\{\rho^{(2)}(z) - \lambda^2\right\} e^{-2\pi i w\cdot z}\, dz \\ &= \lambda + \int_0^\infty \left\{\rho_I^{(2)}(r) - \lambda^2\right\} r^{d-1} \int_{S^{d-1}} e^{-2\pi r i\ w\cdot u}\, du\, dr \\ &= \lambda + 2\pi \int_0^\infty \left\{\rho_I^{(2)}(r) - \lambda^2\right\} \left(\frac{r}{t}\right)^{d/2-1} \mathcal{J}_{d/2-1}(2\pi t r) r dr \\ &\equiv \lambda + 2\pi\ \mathcal{H}_{d/2-1}\left[\left\{\rho_I^{(2)}(r) - \lambda^2\right\} \left(\frac{r}{t}\right)^{d/2-1}\right](2\pi t), \end{aligned} \tag{65}$$

i.e. the sdf is related to the Hankel transform of the complete covariance function.

And finally recall the radially averaged set covariance function,

$$\bar{\nu}_B(r) := \frac{1}{|S^{d-1}|} \int_{S^{d-1}} |B \cap B_{ru}| du, \quad r \geq 0$$

where $\nu_B(z) := |B \cap B_z|$ is the set covariance function of the set B (Illian et al., 2008, Appendix B.3).

To gain insight into how we might estimate the radial sdf besides numerically rotation-averaging the d-dimensional periodogram using Eq. (49), let us analytically rotation-average Bartlett's periodogram (Eq. 20): with $t \geq 0$

$$\begin{aligned} \bar{I}_0(t) &:= \frac{1}{|S^{d-1}|} \int_{S^{d-1}} I_0(tu)\, du \qquad (66) \\ &= \frac{1}{|B||S^{d-1}|} \sum_{x\in X\cap B} \sum_{y\in X\cap B} \int_{S^{d-1}} e^{-2\pi i (tu)\cdot(x-y)}\, du \\ &= \hat{\lambda} + \frac{1}{|B||S^{d-1}|} \sum_{x,y\in X\cap B}^{\neq} \int_{S^{d-1}} e^{-2\pi t i (x-y)\cdot u}\ du \\ &= \hat{\lambda} + \frac{2\pi}{|B||S^{d-1}|t^{d/2-1}} \sum_{x,y\in X\cap B}^{\neq} \mathcal{J}_{d/2-1}(2\pi t\|x-y\|)\ \|x-y\|^{-(d/2-1)} \qquad (67) \end{aligned}$$

which in the planar case simplifies to

$$\overset{d=2}{=}\ \hat{\lambda} + \frac{1}{|B|} \sum_{x,y\in X\cap B}^{\neq} \mathcal{J}_0(2\pi t\|x-y\|).$$

This explains the motivation behind the isotropic estimator discussed by Bartlett (1964) and Diggle et al. (1987). Them authors prefer to normalize by $N(B)$ rather than by $|B|$, corresponding to further dividing (67) by $\hat{\lambda}$.

Like the classical periodogram, the isotropic "shortcut" estimator of Eq.(67) is highly biased near 0, as has been observed by Diggle et al. (1987). The biases are described in the following proposition.

Proposition 7.1. *Expectation of the isotropic Estimator, I. With the definition of the isotropic estimator given by* (67)*,*

$$\mathbf{E}\left\{\bar{I}_0(t)\right\} = \lambda + 2\pi\ \mathcal{H}_{d/2-1}\left[\left\{\rho_I^{(2)}(r) - \lambda^2\right\} \left(\frac{r}{t}\right)^{d/2-1}\ \frac{\bar{\nu}_B(r)}{|B|}\right](2\pi t) + 2\pi\lambda^2\ \mathcal{H}_{d/2-1}\left[\left(\frac{r}{t}\right)^{d/2-1}\ \frac{\bar{\nu}_B(r)}{|B|}\right](2\pi t)$$

Proof. Follows by applying the Campbell's theorem and adding and subtracting

$$2\pi\lambda^2\ \mathcal{H}_{d/2-1}\left[\left(\frac{r}{t}\right)^{d/2-1}\ \frac{\bar{\nu}_B(r)}{|B|}\right](2\pi t) = \lambda^2 \frac{1}{|S^{d-1}||B|} \int_{S^{d-1}} T(B, tu) du.$$

□

If we compare the formulas for the isotropic sdf in Eq.(65) and the expectation in Proposition 7.1, we recognise two sources of bias, just like with the periodogram: A convolution with a function $\bar{\nu}_B/|B|$, coming from the finite observation window like in the periodogram but this time radially averaged, and a centering bias term that is a radially averaged version of the d-dimensional bias. Diggle's truncation-to-local-minimum

construction, see (19), tries to correct for the biases, but among other things, it will not work well for a clustered process as the spectrum near zero will be underestimated.

To work towards a tapered estimator, consider rotation averaging the debiased, tapered estimator of (30), in expectation,

$$\begin{aligned}&\mathbf{E}\frac{1}{|S^{d-1}|}\int_{S^{d-1}}\widetilde{I}_h(tu)\,du \qquad (68)\\ =&\lambda+\frac{2\pi}{|S^{d-1}|t^{d/2-1}}\mathbf{E}\left[\sum_{x,y\in X\cap B}^{\neq}h(x)h(y)\mathcal{J}_{d/2-1}\left(t\|x-y\|\right)\|x-y\|^{-(d/2-1)}\right]-\frac{\lambda^2}{|S^{d-1}|}\int_{S^{d-1}}|H(tu)|^2\,du.\end{aligned}$$

The problem is that a radial function h_I might not exist for which $h_I(||x-y||)=h(x)h(y)$. But if we consider data-tapering the observed differences $\{x-y : x,y\in X\cap B\}$ instead of in B, we can device a taper h_I in $B\oplus B=\bigcup_{x\in B}\{B+x\}$, thus we propose the following isotropic estimator:

$$\bar{I}_h(t):=\hat{\lambda}+\frac{2\pi}{|S^{d-1}|t^{d/2-1}|B|}\sum_{x,y\in X\cap B}^{\neq}h_I(x-y)\mathcal{J}_{d/2-1}(2\pi t\|x-y\|)\;\|x-y\|^{-(d/2-1)}-\widehat{\lambda^2}Bias_I(t) \qquad (69)$$

where the bias correction term consists of the rotation average of both the taper and the set covariance of B,

$$\begin{aligned}Bias_I(t) &= 2\pi\mathcal{H}_{d/2-1}\left[\left(\frac{r}{t}\right)^{d/2-1}\frac{\bar{p}_{B,h}}{|B|}\right](2\pi t),\quad\text{where}\\ \bar{p}_{B,h}(r) &= |S^{d-1}|^{-1}\int_{S^{d-1}}|B\cap B_{ru}|h_I(ru)du,\end{aligned}$$

and depends only on $t=|w|$. We warn the reader that the estimator of (69) is not necessarily positive, but like previous authors (Lii and Masry, 1994; Bandyopadhyay and Lahiri, 2009), we sacrifice this to remove large part of the bias. Additional issue arises from estimation of the required λ^2, as the standard estimator $\hat{\lambda}^2$ is biased with a term depending on $\lambda/|B|$ and a term depending on the unknown second order product density of the underlying process. In our examples we use the slightly less biased $\widehat{\lambda^2}=N(B)[N(B)-1]/|B|^2$ which removes the first order bias. On the positive side, do note that the debiased isotropic estimator without particular tapering (i.e. $h_I\equiv 1$) can be estimated without any tuning parameters.

We may now ask what radial taper to use. Tapers are normally designed for processes observed in discrete time or space. We have discussed separable filters, and now note that the choice of filter needs to match the observational domain. For example it is difficult to match a square observational domain with a radial filter, unless one chooses a compact filter inscribed by the box. There are therefore three considerations 1) demanding the taper be separable in space, 2) isotropic in space, or 3) exactly compact in space. Discrete and isotropic filters have been been determined numerically by Simons and Wang (2011). This will not be an option for us as we need to evaluate the tapers at random locations.

Slepian and co-authors have studied the design of tapers in arbitrary dimensions (Slepian, 1964). Often solutions are in terms of prolate spheroidal wavefunctions. Their discrete analogue is the prolate spheroidal wave sequences, which are similar to the set of Hermite function (Xu et al., 1999). However, the isotropic estimator can be seen as an estimator for the Fourier transform of the (non-stationary) first moment of the difference process $Z=\{z=x-y : x,y\in X\cap B\}$, with the connection $\rho^{(2)}(z)=\lambda_Z(z)/|B\cap B_z|$.

Assuming the observation window is a cuboid $B=\prod_{j=1}^d[0,\ell_j]$, then the difference vector observation window is $B\oplus B=\prod_{j=1}^d[-\ell_j,\ell_j]$, and we can create the d-dimensional taper related to the Hermite functions as the product of the squared exponentials

$$h_I(z)=\prod_{j=1}^d\exp\left[-az_j^2/(2\ell_j)^2\right],\quad z\in\mathbb{R}^d \qquad (70)$$

We see that if $\ell_j=\ell$ for all j, then h_I becomes radial. This is related to circular harmonic decompositions, see (Jacovitti and Neri, 2000). The authors of (Xu et al., 1999) recommend using a scaling factor which we have adjusted for a sampling region of $z_j\in(-\ell_j,\ell_j)$, but which they match to the spheroidal wave functions. We propose setting $a=25$, as this numerically seems to not down–weight too much data, or have a too large jump near the border of $B\oplus B$.

In a sense our above construction is a "fix" as it only works for the 0th Hermite function, and using more than one taper will even if $\ell_j=\ell$ for all j not lead to isotropic functions. This is not unsurprising as the sampling domain will leave an inprint, and capturing this set of information requires using more than the first taper. Also, if we have a spatial sampling that is highly elongated rectangle rather than a square then another solution needs to be chosen.

Finally we may ask ourselves, what happens if we have a point process X which is anisotropic but we still evaluate (18) in 2D for the pattern. Because it incorporates angular averaging we expect the RHS of the expression to only depend on $\|w\|$ as it does.

Proposition 7.2. *Expectation of Diggle's 2D Estimator II. The large area expectation of Diggle's estimator given in* (18) *when B is a cuboidal box when the spectrum is not necessarily isotropic takes the form in terms of $\widetilde{f}(w)$ as defined in* (11) *of*

$$\mathbf{E}\{I_D(w)\} = \lambda + \lambda^2 \int_{\mathbb{R}^d} \widetilde{f}(v) \frac{\delta(\|w\| - \|v\|)}{\|w\|} dv + \int_{\mathbb{R}^d} \frac{\delta(\|w\| - \|v\|)}{\|w\|} |B|^{-1} T(B, w - v) dv + o(1), \tag{71}$$

where $T(B, w)$ is defined in Lemma 4.1.

Proof. The proof is provided in Appendix R. □

8 Simulation studies

To study the behaviour of the estimators we have discussed, and to verify some of the debiasing results, we conducted a simulation study.

8.1 Details for executing the simulations study

We simulated three stationary models in 2D to represent the archetypal spatial patterns of regularity, complete randomness and clustering. Throughout the simulations the intensity, i.e. expected number of points per unit area, was kept constant at $\lambda \equiv 0.01$. The processes were simulated in increasing square observation windows $B_n = [-\ell_n/2, \ell_n/2]^2$ such that on average we observed $\lambda|B_n| = \lambda\ell_n^2 = n = 25, 50, 100, 200, 400$ or 800 points. For example, $\ell_{n=100} = 100$ and $\ell_{n=400} = 200$ spatial units. We chose the regime both to explore the often encountered small data scenarios and to ascertain asymptotic behaviour. The models below are tuned so that they produce patterns with visually distinct structures at $\ell_{n=100}$.

Complete randomness i.e. the Poisson process has no adjustable parameters after λ was fixed. Spatial regularity is represented by the Matérn type II process (Illian et al., 2008, Section 6.5.2) in two variants, one with hard-core radius 2 (variant "r2") and one with radius 5 (variant "r5") which correspond to about 40% and 90% of the maximum allowed radius for the process given λ, respectively. The clustering behaviour is represented by the Thomas process (Illian et al., 2008, Section 6.3.2) with cluster intensity κ and Gaussian dispersal kernel standard deviation σ and again in two variants, one with many small or tight clusters ($\kappa = 0.6\lambda, \sigma = 2$, variant "MS") and one with a few large clusters ($\kappa = 0.3\lambda, \sigma = 6$, variant "FL"). The per-cluster expected point count μ got then fixed via the property $\mu\kappa = \lambda$ to 1.67 and 3.33, respectively. The theoretical pair correlation functions and spectral density functions of the models are depicted in Figure 1. Example patterns of each model and observation window combination are given in Supplement Figure 6.

For the estimation of the 2D spectra we fixed the wavenumber grid to $w \in [-0.3, 0.3]^2$ with 101 equidistant steps in each dimension, giving the step size 0.006 in both dimensions. This scale was chosen to cover the interesting range of non-constant values for all models (cf. Figure 1). To reduce the effect of high wave number noise only the values on the sub-grid $w \in [-0.2, 0.2]^2$ were considered when integrating over w for the quality metrics discussed below. When a rotationally averaged curve was to be computed using Eq. (49), the averaging was done over the magnitude-grid $|w| = 0.003, 0.006, ..., 0.300$ using a box kernel and if not otherwise stated, radius of $1.25 \cdot 0.006$. The radial Hermite taper parameter was fixed to $a = 25$.

We summarised the quality of each estimator $I_\cdot = I_\cdot(w; \mathbf{x})$ under all combinations of a model M and observation window B_n, say M_n, with an integrated summary. First, we estimated per-wavenumber variance $V(w; M_n) = \mathrm{Var}[I_\cdot(w; \mathbf{x})|\mathbf{x} \sim M_n]$, bias $Bias(w; M_n) = \mathbf{E}[I_\cdot(w; \mathbf{x}) - f_M(w)|\mathbf{x} \sim M_n]$ and mean square error $MSE(w; M_n) = V(w; M_n) + Bias^2(w; M_n)$. The f_M stand for the theoretical sdf of the model M. Then these were summarised further to $iVar(M_n) = \sum_w V(w; M_n)$, the integrated squared bias $iBias^2(M_n) = \sum_w Bias^2(w; M_n)$ and $iMSE(M_n) = iVar(M_n) + iBias^2(M_n)$. Smaller $iVar, iBias^2$ and $iMSE$ indicate better quality. The quantities were estimated from 1000 simulations of every M_n.

The estimator for a periodogram is given in equation (30) with the debiasing term included. Bartlett's periodogram ("periodogram") has the taper $h_0(x) = 1_{B_n}(x)/\sqrt{|B_n|}$ for window B_n. For the multitapering ("mt P"), defined in equation (41) with P^2 tapers each having a parameter $p = (p_1, p_2) \in \{1, ..., P\}^2$, we used the orthogonal sine-tapers $h_p(x) = \mathbf{1}_{B_n}(x) \prod_{j=1}^2 \sin[\pi p_j (x_j + l_n/2)/l_n]$. For the kernel smoothed estimators ("smoothed m"), we first compute the Bartlett's periodogram and then, given the estimate $\hat{I}_0$ on the wavenumber gird, we convolute it with a $m \times m$ discrete template having approximately Gaussian weights.

All computations were done using the R-software (R Core Team, 2020, v.3.6.3). Simulations were done with the help the R-package `spatstat` (Baddeley et al., 2015, v.1.64-1). Each of the discussed estimators were programmed into an R-package `ppspectral`, available on request from the first author.

8.2 The effect of debiasing

As we saw in Figure 2, the debiasing term has a large effect on the quality of the estimates. Table 1 provides the fractions of bias removed by the debiasing term for the Bartlett's periodogram and $P = 3$ multitapered periodogram. Nearly all bias is removed, with the only notable exception being the 65% bias left in the multitapered periodogram when data is a (arguably very) small sample of Thomas FL variant.

Model	Estimator	n =25	50	100	200	400	800
MaternII r5	periodogram	1.00	1.00	1.00	0.98	0.99	1.00
MaternII r5	mt P=3	0.99	1.00	1.00	1.00	1.00	1.00
MaternII r2	periodogram	1.00	1.00	1.00	0.99	0.99	1.00
MaternII r2	mt P=3	1.00	1.00	1.00	1.00	1.00	1.00
Poisson	periodogram	1.00	1.00	1.00	0.98	0.99	1.00
Poisson	mt P=3	1.00	1.00	1.00	1.00	1.00	1.00
Thomas FL	periodogram	0.93	0.98	0.99	0.97	0.99	1.00
Thomas FL	mt P=3	0.35	0.86	0.98	1.00	1.00	1.00
Thomas MS	periodogram	0.98	0.99	1.00	0.97	0.99	1.00
Thomas MS	mt P=3	0.89	0.98	1.00	1.00	1.00	1.00

Table 1: The fraction of bias removed by the proposed bias correction. Maximum is 1.00.

We illustrate the form of the biases in Figure 3 for two model variants. The centering bias is always positive, as expected from equation (22). Bartlett's periodogram bias (cf. Lemma 5.1) is concentrated along the axis where one of the sinc-functions is constant 1 and near the origin where the second sinc function grows to 1. The multitapered periodogram has a square-shape around the origin.

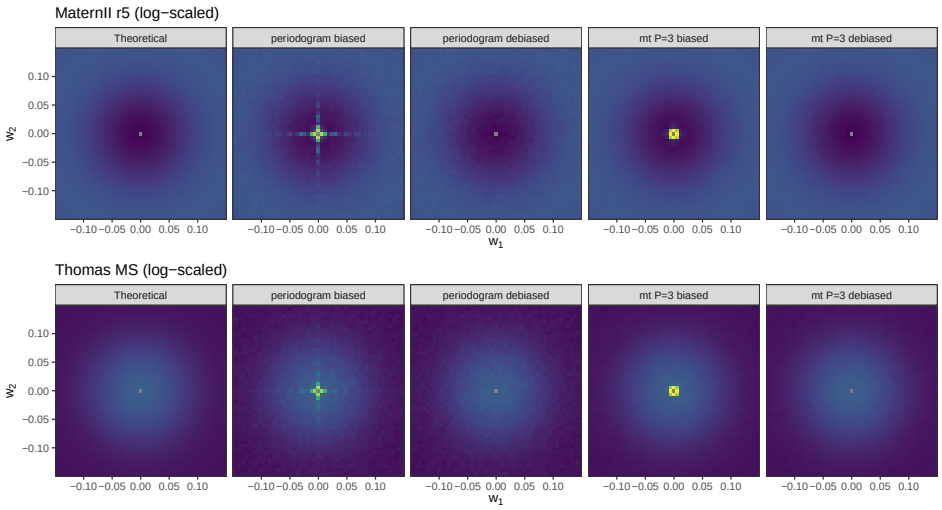

Figure 3: The effect of bias removal for two of the models described in the text. Theoretical sdf on the left, and then towards the right, estimated mean of: Bartlett's periodogram, biased and debiased; Multitapered $P = 3$, biased and debiased. Estimated with sample sizes $n = 800$.

8.3 Overall quality of the estimators

Figure 4 summarizes the estimated integrated variance, squared bias and MSE of the estimators with various parameters. Only the debiased estimators are included, and the values are given in $\log_{10}$-scale and relative to the baseline given by Bartlett's periodogram.

Multitaper with $P = 1$ (i.e. only one taper) does not differ in iMSE from the periodogram as the benefit of averaging does not apply. With more than one taper the variance is reduced, e.g. $P = 3$ (9 tapers) it goes down by 90%, and the bias stays at baseline. But adding more and more tapers ($P = 6$, 36 tapers) starts to introduce bias especially for low sample sizes. The bias is more prominent for the clustered cases, and based on the earlier discussion this is due to oversmoothing at small wavenumbers. Both the variance reduction and the oversmoothing happens with the post-hoc smoothing as well.

From the results we can confirm that some smoothing is beneficial but again a balance must be struck to avoid oversmoothing (cf. Section 6). The regular process is more resilient to oversmoothing with reasonable sample sizes.

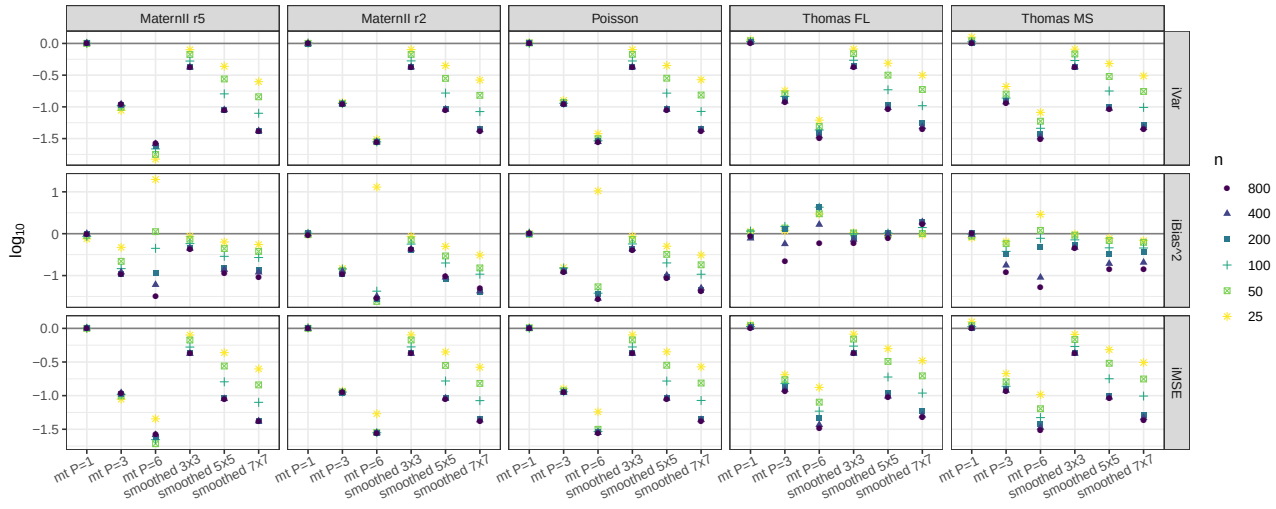

Figure 4: Relative integrated variance (top row), squared integrated bias (middle), and integrate mean square error (bottom) for different models and estimators of spectral density at various levels of sample size n. The values are relative to debiased Bartlett's periodogram, which has the value 0 in the presented $\log_{10}$-scale. For example, a value of -1 means the error is of order 0.1 relative to the periodogram.

8.4 Rotational averaging and isotropic estimation

Isotropic estimation, either by averaging radially or by the isotropic shortcuts discussed in Section 7, is an important dimension reduction step for easier visual inspection of the periodogram (particularly for 3D data). The kernel-based rotation averaging depends on a bandwidth parameter, and we studied how this affects the quality metrics. We set the bandwidth radius to $BF \cdot 0.006$ and varied the bandwidth factor BF from 1 to 10. Figure 5 provides the $iBias^2$ and $iMSE$ for the debiased Bartlett's periodogram when comparing the rotationally averaged estimates to the true 1D isotropic sdfs (cf. Figure 1).

The Poisson process has a constant sdf so oversmoothing will not be penalised. For the clustered processes even a small bandwidth introduces bias, but this is typically more than offset by the reduction in variance. The error for the few large clusters variant is more sensitive to changes in the bandwidth than the many small clusters. This is because the effective range of wavenumbers where the sdf exhibits structure is more concentrated, and thus easier to smoothen out, than the few large cluster variant (cf. Figure 1).

The strongly regular variant "r5" is similar to the clustered cases in that an optimal bandwidth is easy clearly present. For the less regular "r2" variant the error is like that of the Poisson process, possibly due to the target sdf being a weak oscillation around a constant and this being hard to detect. If we look the integrated squared bias we however do see that with large enough data oversmoothing is detectable.

We note that since multitapered periodograms and locally averaged periodograms are already smoothed, their rotational averaged estimates exhibit smoothing bias at smaller bandwidth factors (results not worth showing here).

To compare the radial averaging to direct isotropic estimation, we estimated the debiased isotropic periodogram and the squared-exponential tapered isotropic periodogram with $a = 25$. Table 2 provides the relative quality scores for the isotropic and the tapered isotropic periodograms relative to radially averaged periodogram with the best bandwidth for each model and data size, cf. Figure 5. The values are medians over the sample sizes $n \geq 50$.

In iMSE sense the radially averaged periodogram performs the best, chiefly because iMSE is of order iVar and the radial averaging greatly reduces the variance. However, in terms of bias the isotropic estimator provides more accuracy, except for Poisson when no oversmoothing is possible. Adding a smooth taper to the isotropic estimator further reduces the bias, but can increase the variance, likely due to some data near the edges is being filtered out. The bias occurs near $|w| = 0$, mostly when $|w| < 1/(\sqrt{2}l_n)$ (illustration in Supplements Figure 7).

9 Discussion

The Fourier transform of a point process has been treated by numerous authors (Mugglestone and Renshaw, 1996a; Bartlett, 1964, 1963; Diggle et al., 1987). Despite this fact we have seen limited methodological development, leaving methodological understanding lagging much behind that for time series and random fields. There are a number of barriers for using and understanding the DFT of a point process. The first is computational:

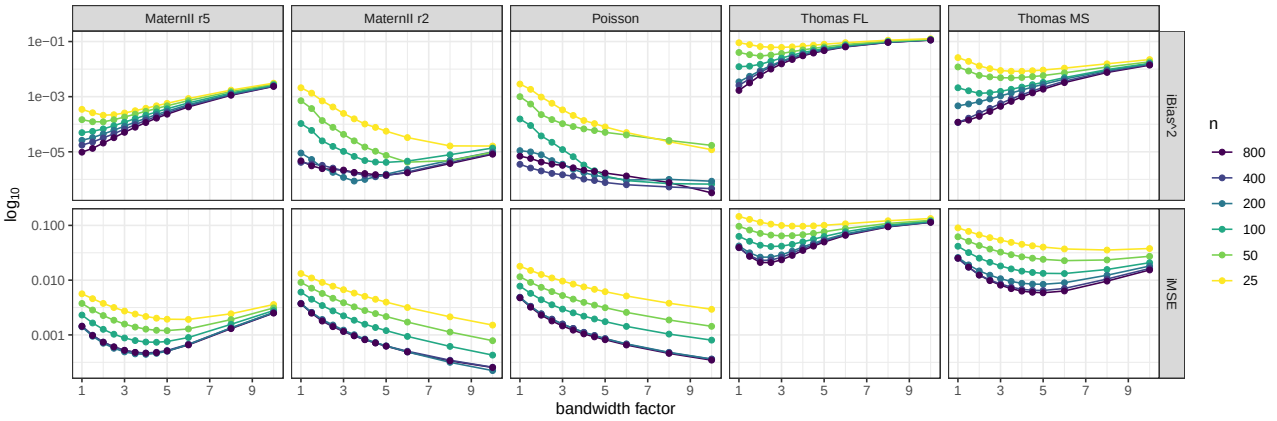

Figure 5: Integrated squared bias (top) and integrated MSE of the rotationally averaged Bartlett's periodogram when compared to the true 1D isotropic sdf. The rotational averaging bandwidth was $BF \cdot 0.006$, where the bandwidth factor BF is on the x-axis. The models are described in the text.

	No taper			Squared-exponential taper		
Model	iMSE	iBias^2	iVar	iMSE	iBias^2	iVar
MaternII r5	6.12	0.21	8.57	3.51	0.29	4.90
MaternII r2	22.40	3.52	23.07	15.59	1.09	16.12
Poisson	15.72	10.53	15.89	13.76	5.75	13.77
Thomas FL	1.22	0.04	1.81	1.32	0.14	1.97
Thomas MS	4.34	0.01	6.50	3.58	0.04	5.35

Table 2: Scores $iMSE, iBias$ and $iVar$, relative to radially averaged debiased periodogram with optimised smoothing bandwidth.

unlike the FFT the act of calculating the DFT is not $\mathcal{O}(N \log N)$, where N are the number of observed points.

Second, there has been an often said, that Fourier analysis of point processes are not very different from that of random fields. However, the results we have developed here, show that the spectral moments are different between random fields and point processes, e.g. compare to (Stein, 2012). Following on from this realisation we have found a way of avoiding the treatment of low wave numbers differently. An important aspect of the spectral moments is the impact of the spatial domain we observe the point process in: the geometry of this object determines what wave numbers we should evaluate the Fourier transform at. Unlike for a regularly sampled stochastic process like a random field, the point process is biased at low wave numbers. This is because, as we demonstrate, we need to de–mean the spectral measure. By understanding the moments of the Fourier transform we show that we may remove this bias completely. Our realisation in this issue is that the sampling domain determines both bias and second order structure for a point process and needs to be studied directly. This has implications for our sampling of the wave number grid, where once we have debiased estimates can be formed at any wave numbers.

Third, we discuss various forms of variance reduction of a spectral variance estimator (the periodogram). We propose either a kernel smoothing of the debiased periodogram, or multitapering combined with a debiasing step. Our studies agree with the theory that unless the second derivative of the spectrum is zero, there is an optimal smoothing width. We also discuss constructing one-dimensional summaries from the two-dimensional spectrum, and tapering in many dimensions.

We have here taken the first steps to fully developing the methodology for spectral analysis of point processes, by calculating the moments of spectral summaries, deducing Gaussianity of the DFT, and fully determining how the wave number grid is influenced by the geometry of the region we observe the point process in. We imagine that authors starting from results such as those provided by this paper will seek to fill in the space remaining in between technical understanding of point processes (Daley and Vere-Jones, 2003) to their practical application (Lavancier et al., 2015; Møller and Torrisi, 2007).

10 Acknowledgements

This work was supported in part by by the UK Engineering and Physical Sciences Research Council under Mathematical Sciences Leadership Fellowship EP/I005250/1, Developing Leaders Award EP/L001519/1, and Award EP/N007336/1; and the European Research Council under Grant CoG 2015-682172NETS within the

Seventh European Union Framework Program.

References

R. Adler and J. E. Taylor. *Random Fields and Geometry.* Springer Verlag, Berlin, 2010.

A. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R.* Chapman and Hall/CRC Press, London, 2015.

S. Bandyopadhyay and S. N. Lahiri. Asymptotic properties of discrete Fourier transforms for spatial data. *Sankhyā: The Indian Journal of Statistics, Series A*, page 221–259, 2009.

M. S. Bartlett. The spectral analysis of point processes (with discussion). *J. Roy. Statist. Soc. Ser. B*, page 264–296, 1963.

M. S. Bartlett. The spectral analysis of two-dimensional point processes. *Biometrika*, 51(3/4):299–311, 1964.

C. A. N. Biscio and R. Waagepetersen. A general central limit theorem and a subsampling variance estimator for α-mixing point processes. *Scandinavian Journal of Statistics*, 46(4):1168–1190, 2019.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer Verlag, Berlin, Germany, 1991.

T. P. Bronez. Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(12):1862–1873, 1988.

S.N. Chiu, D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and Its Applications.* Wiley Series in Probability and Statistics. Wiley, 2013. ISBN 9781118658253.

I.–S. Chou and K.–S. Lii. Spectral analysis of random fields with random sampling. In *Probabilistic and Stochastic Methods in Analysis, with Applications*, page 343–368. Springer, 1992.

E. A. K. Cohen. Multi-wavelet coherence for point processes on the real line. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 2649–2653. IEEE, 2014.

L. Cohen. *Time-frequency analysis*, volume 778. Prentice hall, 1995.

J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

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

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

D. J. Daley. Weakly stationary point processes and random measures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3):406–428, 1971.

D. J Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume I: Elementary Theory and Methods.* Springer Verlag, Berlin, Germany, 2003.

P. Das and B. Babadi. Multitaper spectral analysis of neuronal spiking activity driven by latent stationary processes. *Signal Processing*, 170:107429, 2020.

A. DasGupta. *Asymptotic theory of statistics and probability.* Springer Science & Business Media, 2008.

S. Deb, M. Pourahmadi, and W. B. Wu. An asymptotic theory for spectral analysis of random fields. *Electronic Journal of Statistics*, 11(2):4297–4322, 2017.

P. J. Diggle, D. J. Gates, and A. Stibbard. A nonparametric estimator for pairwise-interaction point processes. *Biometrika*, 74(4):763–770, 1987.

T. Duong and M. L. Hazelton. Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics*, 32(3):485–506, 2005.

D. Durran, J. A. Weyn, and M. Q. Menchaca. Practical considerations for computing dimensional spectra from gridded data. *Monthly Weather Review*, 145(9):3901–3910, 2017.

R. M. Errico. Spectra computed from a limited area grid. *Monthly weather review*, 113(9):1554–1562, 1985.

R. Estrada. On radial functions and distributions and their Fourier transforms. *Journal of Fourier Analysis and Applications*, 20(2):301–320, 2014.

P. F. Fougere. A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis. *Journal of geophysical research*, 82(7):1051–1054, 1977.

L. Grafakos and G. Teschl. On Fourier transforms of radial functions and distributions. *Journal of Fourier Analysis and Applications*, 19(1):167–179, 2013.

A. P. Guillaumin, A. M. Sykulski, S. C. Olhede, and F. J. Simons. Efficient parameter estimation of sampled random fields. *arXiv preprint arXiv:1907.02447*, 2019.

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

A. Hanssen. Multidimensional multitaper spectral estimation. *Signal Processing*, 58(3):327–332, 1997.

H. Hindberg, Y. Birkelund, T. A. Øigård, and A. Hanssen. Kernel-based estimators for the Kirkwood-Rihaczek time-frequency spectrum. In *2006 14th European Signal Processing Conference*, page 1–5, 2006.

J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons Ltd., 2008.

G. Jacovitti and A. Neri. Multiresolution circular harmonic decomposition. *IEEE Transactions on Signal Processing*, 48(11):3242–3247, 2000.

J. Jowett, D. Vere-Jones, and P. A. W. Lewis. The prediction of stationary point processes. In *Stochastic point processes*. Wiley-Interscience, 1972.

J. T. Kent and K. V. Mardia. Spectral and circulant approximations to the likelihood for stationary Gaussian random fields. *Journal of Statistical Planning and Inference*, 50(3):379–394, 1996.

H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty–iii: The dimension of the space of essentially time-and band-limited signals. *Bell Labs Technical Journal*, 41(4): 1295–1336, 1962.

F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(4):853–877, September 2015. ISSN 13697412. doi: 10.1111/rssb.12096.

K.-S. Lii and E. Masry. Spectral estimation of continuous-time stationary processes from random sampling. *Stochastic Processes and Their Applications*, 52(1):39–64, 1994.

T.-C. Liu and B. Van Veen. Multiple window based minimum variance spectrum estimation for multidimensional random fields. *IEEE Transactions on Signal Processing*, 40(3):578–589, 1992.

P. J. Loughlin, J. W. Pitton, and L. E. Atlas. Bilinear time-frequency representations: New insights and properties. *IEEE Transactions on Signal Processing*, 41(2):750–767, 1993.

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

K. S. Miller. *Complex stochastic processes: an introduction to theory and application*. Addison-Wesley, New York, USA, 1974.

M. A. Mugglestone. *Spectral analysis of spatial processes*. PhD thesis, Mathematics, 1990.

M. A. Mugglestone and E. Renshaw. The exploratory analysis of bivariate spatial point patterns using cross-spectra. *Environmetrics*, 7(4):361–377, 1996a.

M. A. Mugglestone and E. Renshaw. A practical guide to the spectral analysis of spatial point processes. *Computational Statistics & Data Analysis*, 21(1):43–65, 1996b.

M. A. Mugglestone and E. Renshaw. Detection of geological lineations on aerial photographs using two-dimensional spectral analysis. *Computers & Geosciences*, 24(8):771–784, 1998.

M. A. Mugglestone and E. Renshaw. Spectral tests of randomness for spatial point patterns. *Environmental and Ecological Statistics*, 8(3):237–251, 2001.

J. Møller and G. L. Torrisi. The pair correlation function of spatial hawkes processes. *Statistics and Probability Letters*, 77(10):995–1003, 2007.

S. C. Olhede and B. Whitcher. Nonparametric tests of structure for high angular resolution diffusion imaging in Q–space. *The Annals of Applied Statistics*, 5(2B):1293–1327, 2011.

B. G. Osgood. *Lectures on the Fourier Transform and Its Applications*, volume 33. American Mathematical Soc., 2019.

D. B. Percival and A. T. Walden. *Spectral Analysis for Physical Applications*. Cambridge University Press, Cambridge, UK, 1993a.

D. B. Percival and A. T. Walden. *Spectral analysis for physical applications*. Cambridge University Press, Cambridge, UK, 1993b.

O. Ponomarenko and Y. Perun. Spectral analysis of some classes of multivariate random fields with isotropic property. *Theory of Stochastic Processes*, 12:142–150, 2006.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

K. S. Riedel. Adaptive smoothing of the log-spectrum with multiple tapering. *IEEE Transactions on Signal Pocessing*, 44(7):1794–1800, 1996.

K. S. Riedel and A. Sidorenko. Minimum bias multiple taper spectral estimation. *IEEE Transactions on Signal Processing*, 43(1):188–195, 1995.

M. Rosenblatt. *Gaussian and non-Gaussian linear time series and random fields*. Springer Science & Business Media, 2012.

P. J. Schreier and L. L. Scharf. Second-order analysis of improper complex random vectors and processes. *IEEE Transactions on Signal Processing*, 51(3):714–725, 2003.

P. J. Schreier and L. L. Scharf. *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. Cambridge University Press, 2010.

F. J. Simons and D. V. Wang. Spatiospectral concentration in the Cartesian plane. *GEM-International Journal on Geomathematics*, 2(1):1–36, 2011.

D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty–iv: extensions to many dimensions; generalized prolate spheroidal functions. *Bell Labs Technical Journal*, 43(6):3009–3057, 1964.

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

M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Verlag, Berlin, Germany, 2012.

Y. Taleb and E. A. K. Cohen. Multiresolution analysis of point processes and statistical thresholding for wavelet-based intensity estimation. *arXiv preprint arXiv:1803.11202*, 2018.

D. J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.

D. J. Thomson. Quadratic-inverse spectrum estimates: applications to palaeoclimatology. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 332(1627): 539–597, 1990.

C. Velasco and P. M. Robinson. Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association*, 95(452):1229–1243, 2000.

S. Vembu. Fourier transformation of the n-dimensional radial delta function. *The Quarterly Journal of Mathematics*, 12(1):165–168, 1961.

D. Vere-Jones. An elementary approach to the spectral theory of stationary random measures. *Stochastic Geometry*, 65:307, 1974.

A. T Walden. A unified view of multitaper multivariate spectral estimation. *Biometrika*, 87(4):767–788, 2000.

M. P. Wand and M. C. Jones. Comparison of smoothing parameterizations in bivariate kernel density estimation. *Journal of the American Statistical Association*, 88(422):520–528, 1993.

M. P. Wand and M. C. Jones. *Kernel smoothing*. CRC Press, 1994.

D. J. Wingham. The reconstruction of a band-limited function and its Fourier transform from a finite number of samples at arbitrary locations by singular value decomposition. *IEEE Transactions on Signal Processing*, 40(3):559–570, 1992.

Y. Xu, S. Haykin, and R. J. Racine. Multiple window time-frequency distribution and coherence of eeg using Slepian sequences and Hermite functions. *IEEE Transactions on Biomedical Engineering*, 46(7):861–866, 1999.

C Yuen. Comments on modern methods for spectrum estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(3):298–299, 1979.

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

A Additional Figures

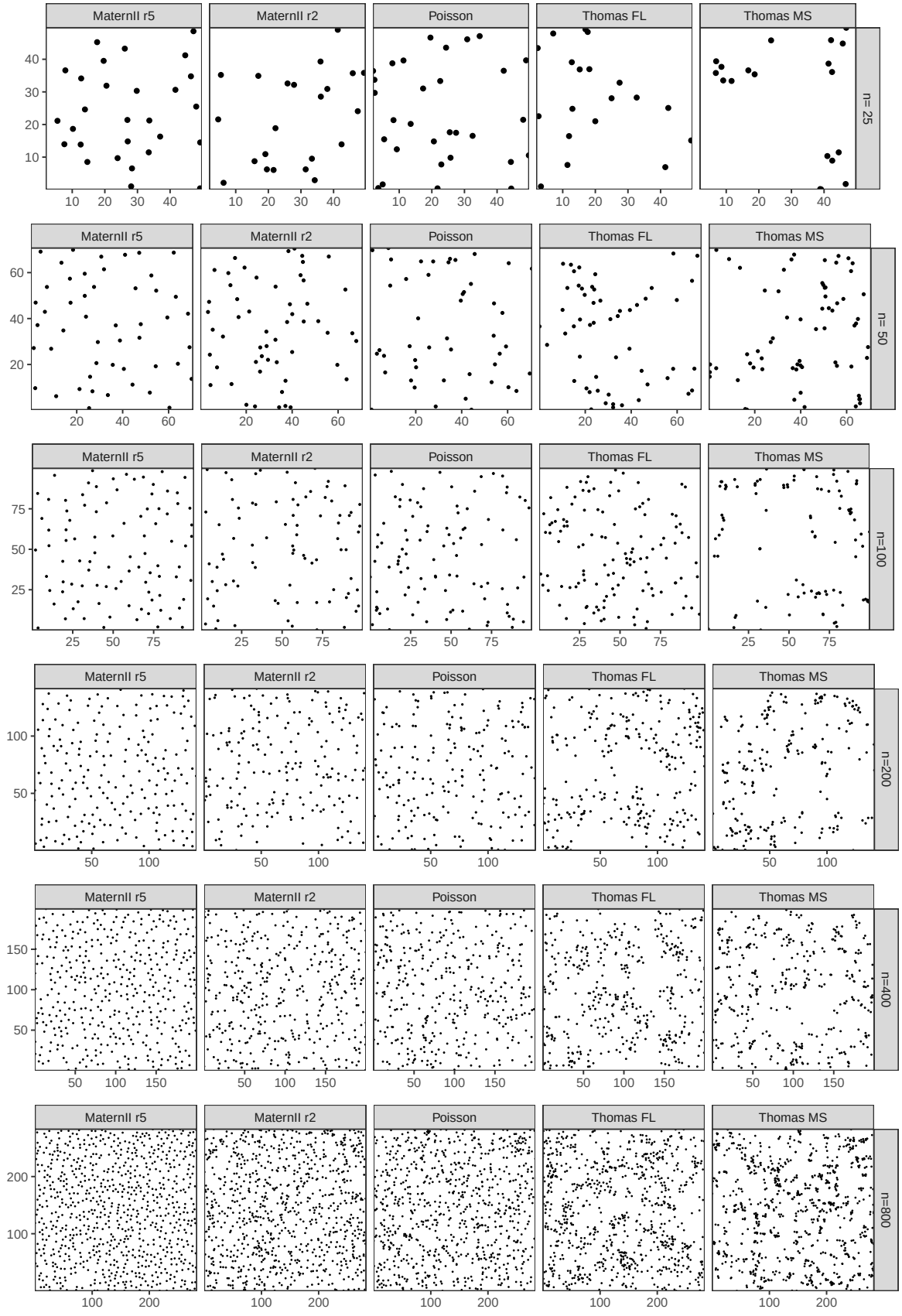

Figure 6: Examples of the simulated patterns used in the simulation trials for studying the quality of the estimators. Details of each model are given in Section 8. Note that the exact number of points in each realisation varies slightly.

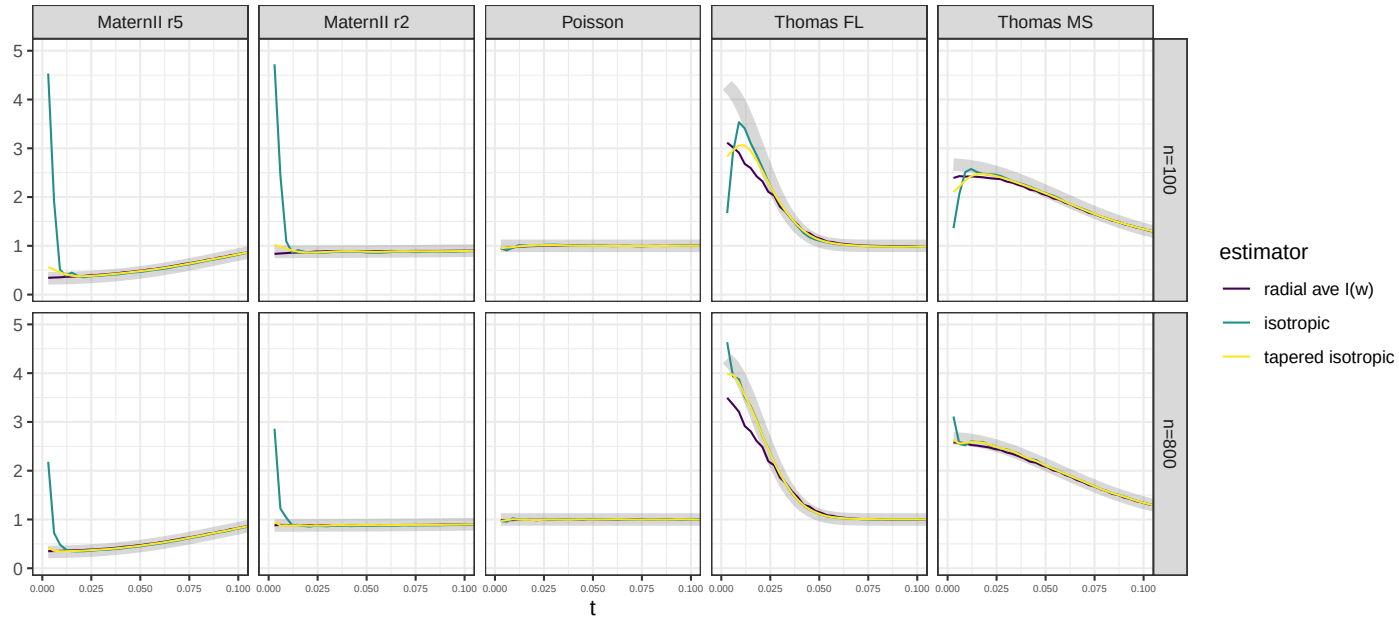

Figure 7: Isotropic sdf/λ, mean estimates versus true curve (thick gray line).

B Proof of Lemma 4.1

Proof. We start the proof by noting that (using that B is centered)

$$\begin{aligned}
\mathbf{E}I_0(w) &= \lambda + |B|^{-1}\int_B\int_{B_{-x}} e^{-2\pi i w\cdot z}\rho^{(2)}(z)dz \qquad (72)\\
&= \lambda + |B|^{-1}\int_B\int_{B_{-x}} e^{-2\pi i w\cdot z}[\rho^{(2)}(z)-\lambda^2]dzdx + |B|^{-1}\lambda^2\int_B\int_{B_{-x}} e^{-2\pi i w\cdot z}dz\\
&= \lambda + |B|^{-1}\int_{\mathbb{R}^d} |B\cap B_{-z}| e^{-2\pi i w\cdot z}[\rho^{(2)}(z)-\lambda^2]dz + |B|^{-1}\lambda^2 T(B,w)\\
&= \lambda + \int_{\mathbb{R}^d} G^{(1)*}(w')\widetilde{G}^{(2)}(w'-w)\,dw' + |B|^{-1}\lambda^2 T(B,w),
\end{aligned}$$

by the convolution theorem, after we have defined the functions:

$$\begin{aligned}
G^{(1)}(w) &= |B|^{-1}\int_{R^d} |B\times B_{-z}| e^{-2\pi i w\cdot z}\,dz = |B|^{-1}T(B,w)\\
\widetilde{G}^{(2)}(w) &= \int_{\mathbb{R}^d} e^{-2\pi i w\cdot z}[\rho^{(2)}(z)-\lambda^2]dz = \lambda^2\widetilde{f}(w).
\end{aligned}$$

Note that from Gradshteyn 3.741 with the change of variables $x_j = \pi w_j \ell_j$ that we get (recalling $|B| = \prod_j \ell_j$):

$$\begin{aligned}
\int_{\mathbb{R}^d} T(B,w)\,dw &= \int_{\mathbb{R}^d}\prod_j \frac{\sin^2(\pi w_j\ell_j)}{(\pi w_j)^2}dw_j\\
&= \int_{\mathbb{R}^d}\prod_j \ell_j^2\frac{\sin^2(x_j)}{x_j^2}(dx_j/(\pi\ell_j))\\
&= |B|\prod_j(\pi/\pi) = |B|. \qquad (73)
\end{aligned}$$

We have defined

$$T(B,w) = \prod_{j=1}^d \frac{\sin^2(\pi w_j\ell_j)}{(\pi w_j)^2}, \quad w\in\mathbb{R}^d. \qquad (74)$$

We now go back to our definition of $\mathbf{E}I_0^{(2)}(w)$, and explore this more carefully. We also fix $0<\beta<1$, and

expand the expectation as follows (realising that care needs to be taken around the pole):

$$\begin{aligned}\mathbf{E}I_0^{(2)}(w) &= \lambda^2 \int_{R^d} G^{(1)*}(w-w')\widetilde{G}^{(2)}(w')dw' \\ &= \lambda^2 \prod_{j=1}^{d} \left(\int_{-\infty}^{w_j - \frac{1}{\ell_j^{1-\beta}}} + \int_{w_j - \frac{1}{\ell_j^{1-\beta}}}^{w_j + \frac{1}{\ell_j^{1-\beta}}} + \int_{w_j + \frac{1}{\ell_j^{1-\beta}}}^{\infty} \right) G^{(1)*}(w-w')\widetilde{G}^{(2)}(w')dw' \\ &= \mathbf{E}I_{1\ldots1}^{(2)}(w) + \mathbf{E}I_{12\ldots}^{(2)}(w) + \cdots + \mathbf{E}I_{d\cdots d}^{(2)}(w). \end{aligned} \tag{75}$$

We can note directly, that taking a supremum over $G^{(1)*}(w-w')$ in the range of integration, and assuming the integral of $G^{(2)}(w)$ is finite (which is true from Parseval-Rayleigh relationships

$$\mathbf{E}I_{11}^{(2)}(w) = \mathcal{O}(|B|^{-1}\prod_{j=1}^{d}\ell_j^{2-2\beta}) = \mathcal{O}(\prod_{j=1}^{d}\ell_j^{1-2\beta}), \tag{76}$$

which requires $\beta > 1/2$ to be $o(1)$. The same holds for the other elements apart from $\mathbf{E}I_{2\ldots2}^{(2)}(k)$. We therefore have

$$\begin{aligned}\mathbf{E}I_0^{(2)}(w) &= \lambda^2 \prod_{j=1}^{d} \int_{w_j - \frac{1}{\ell_j^{1-\beta}}}^{w_j + \frac{1}{\ell_j^{1-\beta}}} G^{(1)*}(w-w')\widetilde{G}^{(2)}(w')dw' \left\{1+o(1)\right\} \\ &\equiv \mathcal{I}_0^{(2)}(w)\left\{1+o(1)\right\}, \end{aligned} \tag{77}$$

so that

$$\mathcal{I}_0^{(2)}(w) \equiv \lambda^2 \prod_{j=1}^{d} \int_{w_j - \frac{1}{\ell_j^{1-\beta}}}^{w_j + \frac{1}{\ell_j^{1-\beta}}} G^{(1)*}(w-w')\widetilde{G}^{(2)}(w')dw'.$$

We assume that $\tilde{f}(w)$ is twice differentiable in w at w (everywhere possibly but $w=0$) so that we get for all $w \neq 0$ and for $\|w-w''\| < \|w-w'\|$

$$\widetilde{f}(w') = \widetilde{f}(w) + \nabla\widetilde{f}(w)^T(w'-w) + \frac{1}{2}(w'-w)^T\tilde{\mathbf{H}}_f(w'')(w'-w), \tag{78}$$

this implicitly defining $\nabla\widetilde{f}(w)$ the gradient, and $\tilde{\mathbf{H}}_f(w)$ as the Hessian matrix.

We can then substitute this back in and need to calculate moments of $G^{(1)}(w)$ to understand the bias. Substituting this expansion back into (77) we get with a change of variables $x_j = \pi\ell_j w_j$:

$$\begin{aligned}\mathcal{I}_0^{(2)}(w) &= \lambda^2 \int_{w_1 - \frac{1}{\ell_1^{1-\beta}}}^{w_1 + \frac{1}{\ell_1^{1-\beta}}} \cdots \int_{w_p - \frac{1}{\ell_p^{1-\beta}}}^{w_p + \frac{1}{\ell_p^{1-\beta}}} G^{(1)*}(w-w') \left\{ \widetilde{f}(w) + \nabla\widetilde{f}(w)^T(w'-w) + \frac{1}{2}(w'-w)^T \right. \\ &\left. \tilde{\mathbf{H}}_f(w'')(w'-w) \right\} dw' \\ &= \lambda^2 \widetilde{f}(w) \int_{w_1 - \frac{1}{\ell_1^{1-\beta}}}^{w_1 + \frac{1}{\ell_1^{1-\beta}}} \cdots \int_{w_p - \frac{1}{\ell_p^{1-\beta}}}^{w_p + \frac{1}{\ell_p^{1-\beta}}} G^{(1)*}(w-w')\,dw' \\ &+ \frac{1}{2}\mathrm{trace}\left\{ \tilde{\mathbf{H}}_f(w'') \int_{w_1 - \frac{1}{\ell^{1-\beta}}}^{w_1 + \frac{1}{\ell^{1-\beta}}} \cdots \int_{w_p - \frac{1}{\ell^{1-\beta}}}^{w_p + \frac{1}{\ell^{1-\beta}}} |B|^{-1} T(B, w-w') \cdot (w'-w)(w'-w)^T \, dw' \right\}. \end{aligned} \tag{79}$$

We can then note that

$$\int_{w_1 - \frac{1}{\ell^{1-\beta}}}^{w_1 + \frac{1}{\ell^{1-\beta}}} \cdots \int_{w_p - \frac{1}{\ell^{1-\beta}}}^{w_p + \frac{1}{\ell^{1-\beta}}} G^{(1)}(w-w')\,dw' = 1 + o(1).$$

By direct calculation we then find

$$\begin{aligned}\int_{R^d} G^{(1)}(w)dw &= |B|^{-1} \int_{R^d}\int_{R^d} |B \times B_{-z}| e^{-i2\pi w\cdot z}\,dzdw \\ &= |B|^{-1} \int_{R^d} |B \times B_{-z}| \delta(z)\,dz \\ &= 1. \end{aligned}$$

Note that this is now consistent with (73), and achieves a value of 1. Second,

$$\int_{w_1-\frac{1}{\ell^{1-\beta}}}^{w_1+\frac{1}{\ell^{1-\beta}}} \cdots \int_{w_p-\frac{1}{\ell^{1-\beta}}}^{w_p+\frac{1}{\ell^{1-\beta}}} (w'-w)^T G^{(1)*}(w-w')\, dw' = 0,$$

as

$$|B \times B_{-z}| = |B \times B_z|,$$

implies that this function $|B \times B_{-z}|$ is symmetric (even) in z, and real-valued, which implies that its Fourier transform also is. The symmetric integral of an even times an odd function is always zero.

We now also determine

$$\mathbf{B} = \int_{w_1-\frac{1}{\ell_1^{1-\beta}}}^{w_1+\frac{1}{\ell_1^{1-\beta}}} \cdots \int_{w_p-\frac{1}{\ell^{1-\beta}}}^{w_p+\frac{1}{\ell^{1-\beta}}} G^{(1)*}(w-w') \cdot (w-w')\,(w-w')^T \; dw'. \tag{80}$$

Let us start by a change of variables. Let us first define the matrix

$$\boldsymbol{L} = \operatorname{diag}(\ell_1, \ldots, \ell_d).$$

Thus the change of variables is

$$\nu = \pi \boldsymbol{L}(w-w') \Longleftrightarrow w' = w - \boldsymbol{L}^{-1}\nu/\pi. \tag{81}$$

With this change of variables we get that

$$\begin{aligned}
\mathbf{B} &= \int_{-\ell_1^\beta/2}^{\ell_1^\beta/2} \cdots \int_{-\ell_p^\beta/2}^{\ell_p^\beta/2} G^{(1)*}(w-w') \cdot (w-w')\,(w-w')^T \; dw' \\
&= \int_{-\ell_1^\beta/2}^{\ell_1^\beta/2} \cdots \int_{-\ell_p^\beta/2}^{\ell_p^\beta/2} \prod_{j=1}^{d} \frac{\pi^2 \ell_j^2 \sin^2(\nu_j)}{|B|\nu_j^2} \boldsymbol{L}^{-1}\pi^{-d}\nu\nu^T\boldsymbol{L}^{-1}\pi^{-d}\, d\nu_j(1/\ell_j) \\
&= \int_{-\ell_1^\beta/2}^{\ell_1^\beta/2} \cdots \int_{-\ell_p^\beta/2}^{\ell_p^\beta/2} \prod_{j=1}^{d} \frac{\ell_j \sin^2(\nu_j)}{|B|\nu_j^2} \boldsymbol{L}^{-1}\nu\nu^T\boldsymbol{L}^{-1}\, d\nu_j.
\end{aligned} \tag{82}$$

We then find if all the $\ell_j = \ell$ then if $0 < \beta$

$$\begin{aligned}
\operatorname{trace}\{\mathbf{B}\} &= \int_{-\ell^\beta/2}^{\ell^\beta/2} \cdots \int_{-\ell^\beta/2}^{\ell^\beta/2} \prod_{j=1}^{d} \frac{\sin^2(\nu_j)}{\nu_j^2} \sum_{q=1}^{d} (\nu_q/\ell)^2 \, d\nu_j \\
&= d \cdot \frac{1}{\ell^2} \left(\int_{-\infty}^{\infty} \frac{\sin^2(x)}{x^2}\, dx \right)^{d-1} \int_{-\ell^\beta/2}^{\ell^\beta/2} \sin^2(x)\, dx \\
&= \frac{d}{\ell^2} \pi^{d-1} \ell^\beta \{1 + o(1)\}.
\end{aligned} \tag{83}$$

This decreases with increasing ℓ as long as $\beta < 2$ (which follows as we have assumed $\beta < 1$), and as the next error term is $O(\ell^{-\beta})$. We therefore want to chose $\beta \to 1$, or for some fixed $\epsilon > 0$ take $\beta = 1 - \epsilon$. Putting all of these components together we get that

$$\begin{aligned}
\mathbf{E} I_0(w) &= \lambda + |B|^{-1} \int_{\mathbb{R}^d} |B \cap B_{-z}| e^{-2\pi i w \cdot z} [\rho^{(2)}(z) - \lambda^2] dz + |B|^{-1}\lambda^2 T(B, w) \\
&= \lambda + \lambda^2 \widetilde{f}(w) + \frac{1}{2} \operatorname{trace} \left\{ \tilde{\mathbf{H}}_f(w'') \frac{d}{\ell^{2-\beta}} \pi^{d-1} \{1 + o(1)\} \right\} + |B|^{-1}\lambda^2 T(B, w) \\
&= \lambda + (-\lambda + f(w)) \{1 + o(1)\} + \frac{1}{2} \operatorname{trace} \left\{ \tilde{\mathbf{H}}_f(w'') \frac{d}{\ell^{2-\beta}} \pi^{d-1} \{1 + o(1)\} \right\} + |B|^{-1}\lambda^2 T(B, w) \\
&= f(w) \{1 + o(1)\} + o(1) + |B|^{-1}\lambda^2 T(B, w).
\end{aligned}$$

□

C Proof of Proposition 4.1

Proof. We start from using Eqn. (9) in Section 2 and determine

$$\begin{aligned}\mathbf{E}\left\{J_h(w)\right\} &= \mathbf{E}\sum_{x\in X} h(x)\exp(-2\pi i w\cdot x)\\ &= \lambda\int_{R^d} h(x)\exp(-2\pi i w\cdot x)\,dx\\ &= \lambda H(w).\end{aligned} \tag{84}$$

Subsequently we find using Eqn. (9) in Section 2

$$\begin{aligned}\mathbf{E}\left\{\left|J_h(w)\right|^2\right\} &= \mathbf{E}\left\{\sum_{x\in X}\sum_{y\in X} h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\right\}\\ &= \mathbf{E}\sum_{x\in X}\left|h(x)\right|^2 + \mathbf{E}\left\{\sum_{x\in X,y\in X}^{\neq} h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\right\}\\ &= \lambda\int_{R^d}\left|h(x)\right|^2 dx + \iint_{R^d\times R^d}\bar{\rho}^{(2)}(x-y)h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\,dx\,dy\\ &= \lambda + \iint_{R^d\times R^d}\bar{\rho}^{(2)}(x-y)h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\,dx\,dy.\end{aligned} \tag{85}$$

We now calculate the variance as

$$\begin{aligned}\mathbf{Var}\left\{J_h(w)\right\} &= \mathbf{E}\left\{\left|J_h(w)\right|^2\right\} - \left|\mathbf{E}\left\{J_h(w)\right\}\right|^2\\ &= \lambda + \iint_{R^d\times R^d}\bar{\rho}^{(2)}(x-y)h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\,dx\,dy - \left|\lambda H(w)\right|^2.\end{aligned} \tag{86}$$

We now define the renormalized quantity

$$\widetilde{\rho}^{(2)}(z) = \frac{\rho^{(2)}(z)-\lambda^2}{\lambda^2},\quad z\in\mathbb{R}^d. \tag{87}$$

The expression in (86) then can be simplified to

$$\begin{aligned}\mathbf{Var}\left\{J_h(w)\right\} &= \lambda + \iint_{R^d\times R^d}\rho^{(2)}(x-y)h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\,dx\,dy - \left|\lambda H(w)\right|^2\\ &= \lambda + \lambda^2\iint_{R^d\times R^d}\left(\widetilde{\rho}^{(2)}(x-y)+1\right)h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\,dx\,dy - \left|\lambda H(w)\right|^2\\ &= \lambda + \lambda^2\iint_{R^d\times R^d}\widetilde{\rho}^{(2)}(x-y)h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\,dx\,dy\\ &= \lambda + \lambda^2\mathbf{Var}_1\left\{J_h(w)\right\},\end{aligned} \tag{88}$$

where we define

$$\mathbf{Var}_1\left\{J_h(w)\right\} \equiv \iint_{R^d\times R^d}\widetilde{\rho}^{(2)}(x-y)h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\,dx\,dy. \tag{89}$$

To simplify this expression we note that

$$\begin{aligned}\mathbf{Var}_1\left\{J_h(w)\right\} &= \int_{R^d\times R^d}\widetilde{\rho}^{(2)}(x-y)h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\,dx\,dy\\ &= \int_{R^d\times R^d}\int_{R^d}\widetilde{f}^{(2)}(k')e^{2\pi i(x-y\cdot w'}dw'\cdot h(x)\exp(-2\pi i w\cdot x)h^*(y)\exp(2\pi i w\cdot y)\,dx\,dy\\ &= \int_{R^d\times R^d}\int_{R^d}\widetilde{f}^{(2)}(k')e^{-2\pi i(x-y)\cdot(w-w')}dw'\cdot h(x)h^*(y)\,dx\,dy\\ &= \int_{R^d}\widetilde{f}^{(2)}(w')\left|H(w-w')\right|^2 dw'.\end{aligned} \tag{90}$$

This yields the desired expression for the variance. We return to the expression of the relation of the DFT. We aim to show that

$$\mathrm{Rel}\left\{J_h(w)\right\} = \lambda \int_{\mathbb{R}^d} H(w'-2w)H(w')dw' + \int_{\mathbb{R}^d} U(w,z)e^{-2\pi iw\cdot z}\{\rho(z)-\lambda^2\}dxdz.$$

We have

$$\mathrm{Rel}\left\{J_h(w)\right\} = \mathbf{E}\left\{\left(J_h(w)\right)^2\right\} - \left(\mathbf{E}\left\{J_h(w)\right\}\right)^2.$$

We then note that

$$\begin{aligned}
\mathbf{E}\left\{\left(J_h(w)\right)^2\right\} &= \mathbf{E}\left\{\sum_{x\in X}\sum_{y\in X} h(x)\exp(-2\pi iw\cdot x)h(y)\exp(-2\pi iw\cdot y)\right\} \\
&= \mathbf{E}\sum_{x\in X}\left\{h(x)\right\}^2 e^{-4\pi iw\cdot x} + \mathbf{E}\left\{\sum_{x\in X,y\in X}^{\neq} h(x)e^{-2\pi iw\cdot x}h(y)e^{-2\pi iw\cdot y}\right\} \\
&= \lambda\int_{R^d}\left\{h(x)\right\}^2 e^{-4\pi iw\cdot x}dx + \iint_{R^d\times R^d}\bar{\rho}^{(2)}(x-y)h(x)e^{-2\pi iw\cdot x}h(y)e^{-2\pi iw\cdot y}\,dx\,dy \\
&= \lambda\int_{R^d}\left\{h(x)\right\}^2 e^{-4\pi iw\cdot x}dx + \mathrm{Rel}_1\left\{J_h(w)\right\}. \qquad (91)
\end{aligned}$$

We now seek to simplify this and write

$$\mathrm{Rel}_1\left\{J_h(w)\right\} = \lambda^2\iint_{R^d\times R^d}\left(\widetilde{\rho}^{(2)}(x-y)+1\right)h(x)e^{-2\pi iw\cdot x}h(y)e^{-2\pi iw\cdot y}\,dx\,dy. \qquad (92)$$

We additionally have from (84):

$$\mathbf{E}^2\left\{J_h(w)\right\} = \lambda^2 H^2(w).$$

Putting all of this together we get that

$$\begin{aligned}
\mathrm{Rel}\left\{J_h(w)\right\} &= \lambda\int_{R^d}\left\{h(x)\right\}^2 e^{-4\pi iw\cdot x}dx + \lambda^2\iint_{R^d\times R^d}\left(\widetilde{\rho}^{(2)}(x-y)+1\right)h(x)e^{-2\pi iw\cdot x}h(y)e^{-2\pi iw\cdot y}\,dx\,dy \\
&\quad - \lambda^2 H^2(w) \\
&= \lambda\int_{R^d}\left\{h(x)\right\}^2 e^{-4\pi iw\cdot x}dx + \lambda^2\iint_{R^d\times R^d}\widetilde{\rho}^{(2)}(x-y)h(x)e^{-2\pi iw\cdot x}h(y)e^{-2\pi iw\cdot y}\,dx\,dy.
\end{aligned}$$

We now implement a change of variables, and set $z = x - y$. We then have

$$\iint_{R^d\times R^d}\widetilde{\rho}^{(2)}(x-y)h(x)e^{-2\pi iw\cdot x}h(y)e^{-2\pi iw\cdot y}\,dx\,dy = \iint_{R^d\times R^d}\widetilde{\rho}^{(2)}(z)h(x)e^{-2\pi iw\cdot x}h(x-z)e^{-2\pi iw\cdot(x-z)}\,dx\,dz,$$

and so with the definition of

$$U(k,z) \equiv \int h(x)h(z+x)e^{-2\pi iw\cdot(2x)}dx, \qquad (93)$$

the expression is a consequence of the above expressions. □

D Proof of Lemma 4.2

Proof. We start from

$$\widetilde{I}_h(w) = \left|J_h(w) - \lambda H(w)\right|^2,$$

and so take

$$\mathbf{E}\left\{\widetilde{I}_h(w)\right\} = \mathbf{E}\{|J_h(w)|^2\} - \lambda\mathbf{E}\{J_h(w)H^*(w)\} - \lambda\mathbf{E}\{J_h^*(w)H(w)\} + \lambda^2\mathbf{E}\{|H(w)|^2\}.$$

We now use Campbell's theorem to evaluate these expectations. We already have from Proposition 4.1 that

$$\mathbf{E}\left\{|J_h(w)|^2\right\} = \lambda + \iint_{R^d\times R^d}\bar{\rho}^{(2)}(x-y)h(x)\exp(-2\pi iw\cdot x)h^*(y)\exp(2\pi iw\cdot y)\,dx\,dy.$$

Also we note that

$$H(w) = \int_{R^d} h(x) \exp(-2\pi i w \cdot x)\, dx. \tag{94}$$

We also note

$$\mathbf{E}\left\{J_h(w)\right\} = \lambda H(w).$$

Therefore we have

$$\begin{aligned}
\mathbf{E}\left\{\widetilde{I}_h(w)\right\} &= \mathbf{E}\{|J_h(w)|^2\} - \lambda \mathbf{E}\{J_h(w)H^*(w)\} - \lambda \mathbf{E}\{J_h^*(w)H(w)\} + \lambda^2 \mathbf{E}\{|H(w)|^2\} \\
&= \mathbf{E}\{|J_h(w)|^2\} - \lambda^2 |H(w)|^2 \\
&= \lambda + \iint_{R^d \times R^d} \{\bar{\rho}^{(2)}(x-y) - \lambda^2\} h(x) \exp(-2\pi i w \cdot x) h^*(y) \exp(2\pi i w \cdot y)\, dx\, dy.
\end{aligned} \tag{95}$$

We additionally note

$$\int_{R^d} \{\bar{\rho}^{(2)}(z) - \lambda^2\} \exp(-2\pi i w \cdot z)\, dz = f(w) - \lambda.$$

This can be noted from calculations in Section 2. We note that the multiplication in (95) leads to a convolution in the wave number domain. If we rewrite

$$\bar{\rho}^{(2)}(z) - \lambda^2 = \int_{R^d} (f(w) - \lambda) \exp(-2\pi i w \cdot z)\, dw, \tag{96}$$

then rewriting (95) we get

$$\begin{aligned}
\mathbf{E}\left\{\widetilde{I}_h(w)\right\} &= \lambda + \iint_{R^d \times R^d} \int_{R^d} (f(u) - \lambda) \exp(-2\pi i u^T (x-y) - 2\pi i w \cdot (x-y))\, du h(x) h^*(y)\, dx\, dy \\
&= \lambda + \int_{R^d} (f(u) - \lambda) \|H(w-u)\|^2\, du = \int_{R^d} f(u) \|H(w-u)\|^2\, du,
\end{aligned}$$

as given. □

E Proof of Theorem 4.1

Proof. The expectation of the tapered Fourier transform is (see (9))

$$\begin{aligned}
\mathbf{E}I^t(w) &= \int_{\mathbb{R}^{2d}} h(x)h(y) e^{-2\pi i w \cdot (x-y)} \rho^{(2)}(x,y)\, dx\, dy + \int_{\mathbb{R}^d} h^2(x) \rho^{(1)}(x)\, dx \\
&= \lambda \int_{\mathbb{R}} h^2(x) dx + \int_{B^2} h(x)h(y) e^{-2\pi i w \cdot (x-y)} \rho^{(2)}(x-y) dx dy \qquad (97) \\
&= \lambda \cdot 1 + \int_{B^2} h(x)h(y) e^{-2\pi i w \cdot (x-y)} [\rho^{(2)}(x-y) - \lambda^2] dx dy \\
&+ \lambda^2 \int_{B^2} h(x)h(y) e^{-i2\pi w \cdot (x-y)} dx dy \\
&= \lambda + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x)h(y) e^{-2\pi i w \cdot (x-y)} [\rho^{(2)}(x-y) - \lambda^2] dx dy + \lambda^2 h(B,w) h(B,-w) \\
&= \lambda + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x)h(y) e^{-2\pi i w \cdot (x-y)} [\rho^{(2)}(x-y) - \lambda^2] dx dy + \lambda^2 |H(w)|^2 \\
&= \lambda + \mathbf{E}I^{t,(2)}(w) + \lambda^2 |H(w)|^2, \qquad (98)
\end{aligned}$$

this defining the quantity $\mathbf{E}I^{t,(2)}(w)$.

The expression $\mathbf{E}I^{t,(2)}(w)$ mirrors what we can see in (72). We see directly the benefits of tapering–$|H(w)|$ will decay rapidly from zero so there will be no effect from the third term, and there will be less blurring in the second term apart from very close to $w = 0$.

From which we see that we can re-express this understanding in the Fourier domain, again using the convolution theorem. We define

$$\mathbf{E}I^{t,(2)}(w) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x)h(y) e^{-2\pi i w \cdot (x-y)} [\rho^{(2)}(x-y) - \lambda^2] dx dy.$$

To explore this further we again do a change of variables:

$$\mathbf{E}I^{t,(2)}(w) = \int_{\mathbb{R}^d} \int_{B_z} h(x)h(x-z)e^{-i2\pi w\cdot z}[\rho^{(2)}(z) - \lambda^2]dxdz.$$

We define the inner product window using that $h(x)$ is compactly supported to get

$$W_h(z) = \int_{B_z} h(x)h(x-z)dx = \int_{\mathbb{R}^d} h(x)h(x-z)dx = \int_{\mathbb{R}^d} |H(\omega)|^2 e^{2\pi i w\cdot z}\, dw.$$

We could also divide by $|B \cap B_{-z}|h(x)h(x-z)$ as long as we are inside the domain. Outside B we cannot, and so this is why we cannot remove all bias in the spectrum even if we bias–correct as suggested by Guyon (1982).

With this window we can note that

$$\mathbf{E}I^{t,(2)}(w) = \lambda^2 \int_{\mathbb{R}^d} |H(w'-w)|^2 \widetilde{f}(w')\, dw'. \tag{99}$$

Ideally we would like $|H(w)|^2 = \delta(w)$, but this is not possible as $|B|$ is finite. If we assume $\widetilde{f}(w)$ is smooth and possessing two derivatives then if B is growing we can obtain nearly unbiasedness. We now assume that $|H(w)|^2$ is concentrated to a region in wave number $\mathcal{W}$ and again use a Taylor expansion. There are two approaches to this, either Riedel (1996) or Thomson (1990).

The systematic bias: $\lambda^2 h(B,w)h(B,-w) = |H(w)|^2$ replaces the sinc functions for the square box. As we have chosen the function h to be well–concentrated Thomson (1990); Riedel and Sidorenko (1995); Riedel (1996); Walden (2000), the effect of this is limited.

We assume that

$$\int_{\mathcal{W}} |H(w)|^2\ dw = 1 - \varepsilon_\ell, \quad \text{where} \quad \varepsilon_\ell = o(1). \tag{100}$$

We also assume that $\tilde{f}(w)$ is upper bounded by $\|\widetilde{f}\|_0$. We then have

$$\mathbf{E}I^{t,(2)}(w) = \lambda^2 \int_{\mathbb{R}^d \setminus \mathcal{W}} |H(w'-w)|^2 \widetilde{f}(w')\, dw' + \lambda^2 \int_{\mathcal{W}} |H(w'-w)|^2 \widetilde{f}(w')\, dw'. \tag{101}$$

We note that

$$\int_{\mathbb{R}^d \setminus \mathcal{W}} |H(w'-w)|^2 \widetilde{f}(w')\, dw' \le \|\widetilde{f}\|_0 \{1 - \{1 - \varepsilon_\ell\}\}.$$

We can yet again utilize the Taylor expansion of (78) inside $\mathcal{W}$. We find

$$\begin{aligned}
\mathbf{E}I^{t,(2)}(w) &= \lambda^2 \int_{\mathcal{W}} |H(w'-w)|^2 \widetilde{f}(w')\, dw' \{1 + o(1)\} \\
&= \lambda^2 \int_{\mathcal{W}} |H(w'-w)|^2 \left\{ \widetilde{f}(w) + \nabla \widetilde{f}(w)^T (w'-w) + \frac{1}{2}(w'-w)^T \tilde{\mathbf{H}}_f(w'')(w'-w) \right\} dw' \{1 + o(1)\} \\
&= \lambda^2 \widetilde{f}(w) + 0 + \lambda^2 \int_{\mathcal{W}} |H(w'-w)|^2 \frac{1}{2}(w'-w)^T \tilde{\mathbf{H}}_f(w'')(w'-w)\ dw'.
\end{aligned} \tag{102}$$

We note that

$$\begin{aligned}
&\| \int_{\mathcal{W}} |H(w'-w)|^2 \frac{1}{2}(w'-w)^T \tilde{\mathbf{H}}_f(w'')(w'-w)\ dw' \|^2 \\
&\le |\tilde{\mathbf{H}}_f(w'')| \int_{\mathcal{W}} |H(w'-w)|^2 \frac{1}{2}(w'-w)^T (w'-w)\ dk' \\
&= |\tilde{\mathbf{H}}_f(w'')| \int_{\mathcal{W}} |H(w)|^2 w^2 dw = \mathcal{O}(1/\ell).
\end{aligned}$$

Then

$$\begin{aligned}
\mathbf{E}I^t(w) &= \lambda + \mathbf{E}I^{t,(2)}(w) + \lambda^2 |H(w)|^2 \\
&= f(k) + 0 + \mathcal{O}(1/\ell) + \lambda^2 |H(w)|^2 .
\end{aligned}$$

We have that $|H(w)|^2$ decays in $|w|$. If $w \notin \mathcal{W}$ then the influence of this term becomes negligible. □

F Proof of Corollary 4.2

Proof. We start from

$$\widetilde{I}_h(w) = \left|\mathrm{DFT}_h\{X\}(w) - \lambda H(w)\right|^2.$$

We calculate the expectation of this object. This takes the form of

$$\mathbf{E}\left\{\widetilde{I}_h(w)\right\} = \mathbf{E}\left|\mathrm{DFT}_h\{X\}(w)\right|^2 - \lambda\mathbf{E}\left\{\mathrm{DFT}_h\{X\}(w)H^*(w) + \mathrm{DFT}_h^*\{X\}(w)H(w)\right\} + \lambda^2\left|H(w)\right|^2. \quad (103)$$

We now use Theorem 4.1 to determine

$$\mathbf{E}I^t(w) = f(w) + \mathcal{O}(1/\ell) + \lambda^2\left|H(w)\right|^2.$$

We now need to determine the middle term of (103).

We need to determine (using Campbell's theorem)

$$\begin{aligned}\mathbf{E}\left\{\mathrm{DFT}_h\{X\}(w)H^*(w)\right\} &= \mathbf{E}\left\{\sum_{x\in\mathcal{X}} h(x)e^{-2\pi iw\cdot x}\right\}H^*(w)\\ &= \lambda H^*(w)\int h(x)e^{-2\pi iw\cdot x}dx\\ &= \lambda\left|H(w)\right|^2. \end{aligned} \quad (104)$$

Also note

$$\mathbf{E}\left\{\mathrm{DFT}_h^*\{X\}(w)H(w)\right\} = \mathbf{E}^*\left\{\mathrm{DFT}_h\{X\}(w)H^*(w)\right\} = \lambda\left|H(w)\right|^2. \quad (105)$$

Therefore we find that

$$\begin{aligned}\mathbf{E}\left\{\widetilde{I}_h(w)\right\} &= f(w) + \mathcal{O}(1/\ell) + \lambda^2\left|H(w)\right|^2 - 2\lambda^2\left|H(w)\right|^2 + \lambda^2\left|H(w)\right|^2\\ &= f(w) + \mathcal{O}(1/\ell). \end{aligned} \quad (106)$$

□

G Proof of Theorem 4.1

Proof. The proof of this result is somewhat more challenging. Assume that the domain B is symmetric with respect to the axis, so that the centre of mass of the region is at the origin. Assume the regions have a positive Euler characteristic, or that there are no holes in the observation region. Determine the mean length in each dimension and overlay a box $\mathcal{B}_\square$. Assume this box has width ℓ_i in dimension i.

Let us first describe this set-up in 2 dimensions and then what modifications are needed in higher dimensions. Now imagine B, the two dimensional observational domain need δB_l removed and δB_u added ti make up $\mathcal{B}_\square$. Thus

$$B \cup \delta B_l = \mathcal{B}_\square \cup \delta B_u.$$

Note that we assume

$$B \cap \delta B_l = \emptyset, \quad \text{and}$$

and also assume that

$$\ell_i = \Theta(\ell),$$

to slave all dimensions to one scale.

It follows that we can construct 3 different cuboids (boxes), one inscribed by the shape, one with the average distance to the origin the same as with the shape (almost overlapping the shape), and one which inscribes the shape. We shall write these cuboids as $\mathcal{B}_0(\ell)$, $\mathcal{B}_1(\ell)$ and $\mathcal{B}_2(\ell)$. We note that the expectation of the modulus square of the Fourier transform is to leading order the same between the three regions, as long as they grow according to the order ℓ^d.

This allows us to treat any region, as long as it is not too skinny or asymmetrical. This is why we slave the growth in any dimension to ℓ. □

H Proof of Theorem 5.1

Proof. From first principles using Campbell's theorem we may deduce that the uncentered second moment of $I_p^t(w_1)$ with $I_l^t(w_2)$ takes the form of

$$\begin{aligned}
\mathbf{E}\left\{I_p^t(w_1)I_l^t(w_2)\right\} =&\mathbf{E}\left[\sum_{x,y}^{\neq} h_p(x)h_p(y)\phi_{w_1}(x-y)+\sum_x h_p^2(x)\right]\left[\sum_{z,v}^{\neq} h_l(z)h_l(v)\phi_{w_2}(z-v)+\sum_q h_l^2(q)\right]\\
=&\mathbf{E}\left\{\sum_{x,y}^{\neq} h_p(x)h_p(y)\phi_{w_1}(x-y)\sum_{z,v}^{\neq} h_l(z)h_l(v)\phi_{w_2}(z-v)\right\}\\
&+\mathbf{E}\left\{\sum_x h_p^2(x)\sum_{y,z}^{\neq} h_l(y)h_l(z)\phi_{w_2}(y-z)+\sum_x h_l^2(x)\sum_{y,z}^{\neq} h_p(y)h_p(z)\phi_{w_1}(y-z)\right\}\\
&+\mathbf{E}\left[\sum_x h_p^2(x)\sum_y h_l^2(y)\right]\\
=&V_1(w_1,w_2)+V_2(w_1,w_2)+V_3. \qquad (107)
\end{aligned}$$

This latter equation defines the three terms $V_1(w_1,w_2)$, $V_2(w_1,w_2)$ and V_3 (where the latter does not depend on w_1 and w_2 even if V_2 is the sum of two terms, one only depending on w_1 and the other only depending on w_2). We now need to further split these terms up in order to learn about the cases when we can have repeats of locations or not. We start with V_1. First, we note:

$$\begin{aligned}
V_1(w_1,w_2) =& \mathbf{E}\left\{\sum_{x,y}^{\neq} h_p(x)h_p(y)\phi_{w_1}(x-y)\sum_{z,v}^{\neq} h_l(z)h_l(v)\phi_{w_2}(z-v)\right\}\\
=&\mathbf{E}\sum_{x,y}^{\neq}\sum_{z,v}^{\neq}\mathbf{1}(x\neq z,x\neq v)\,\mathbf{1}(y\neq z,y\neq v)\,h_p(x)h_p(y)h_l(z)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(z-v)\\
+&\mathbf{E}\sum_{x,y}^{\neq}\sum_{z,v}^{\neq}\mathbf{1}(x=z,x\neq v)\,\mathbf{1}(y\neq z,y\neq v)\,h_p(x)h_p(y)h_l(z)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(x-v)\\
+&\mathbf{E}\sum_{x,y}^{\neq}\sum_{z,v}^{\neq}\mathbf{1}(x\neq z,x=v)\,\mathbf{1}(y\neq z,y\neq v)\,h_p(x)h_p(y)h_l(z)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(z-v)\\
+&\mathbf{E}\sum_{x,y}^{\neq}\sum_{z,v}^{\neq}\mathbf{1}(x\neq z,x\neq v)\,\mathbf{1}(y=z,y\neq v)\,h_p(x)h_p(y)h_l(z)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(z-v)\\
+&\mathbf{E}\sum_{x,y}^{\neq}\sum_{z,v}^{\neq}\mathbf{1}(x\neq z,x\neq v)\,\mathbf{1}(y\neq z,y=v)\,h_p(x)h_p(y)h_l(z)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(z-v)\\
+&\mathbf{E}\sum_{x,y}^{\neq}\sum_{z,v}^{\neq}\mathbf{1}(x=z,x\neq v)\,\mathbf{1}(y=v,y\neq z)\,h_p(x)h_p(y)h_l(z)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(z-v)\\
+&\mathbf{E}\sum_{x,y}^{\neq}\sum_{z,v}^{\neq}\mathbf{1}(x=v,x\neq z)\,\mathbf{1}(y=z,y\neq v)\,h_p(x)h_p(y)h_l(z)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(z-v).
\end{aligned}$$

Using the simplification implied by (9) we obtain the expression of

$$
\begin{aligned}
V_1(w_1,w_2) &= \mathbf{E}\left\{\sum_{x,y}^{\neq} h_p(x)h_p(y)\phi_{w_1}(x-y)\sum_{z,v}^{\neq} h_l(z)h_l(v)\phi_{w_2}(z-v)\right\}\\
&= \int_{B^4}\rho^{(4)}(x,y,z,v)h_p(x)h_p(y)h_l(z)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(z-v)dxdydvdz\\
&+ \int_{B^3}\rho^{(3)}(x,y,v)h_p(x)h_p(y)h_l(x)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(x-v)dxdydv\\
&+ \int_{B^3}\rho^{(3)}(x,y,z)h_p(x)h_p(y)h_l(z)h_l(x)\phi_{w_1}(x-y)\phi_{w_2}(z-x)dxdydz\\
&+ \int_{B^3}\rho^{(3)}(x,y,v)h_p(x)h_p(y)h_l(y)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(y-v)dxdydv\\
&+ \int_{B^3}\rho^{(3)}(x,y,z)h_p(x)h_p(y)h_l(z)h_l(y)\phi_{w_1}(x-y)\phi_{w_2}(z-y)dxdydz\\
&+ \int_{B^2}\rho^{(2)}(x,y)h_p(x)h_p(y)h_l(x)h_l(y)\phi_{w_1}(x-y)\phi_{w_2}(x-y)dxdy\\
&+ \int_{B^2}\rho^{(2)}(x,y)h_p(x)h_p(y)h_l(y)h_l(x)\phi_{w_1}(x-y)\phi_{w_2}(y-x)dxdy.
\end{aligned}
$$

We will be able to use the orthogonality between the data tapers $\{h_l(x)\}$ but this will be easier in the Fourier domain where we can assume local smoothness of the Fourier transform of $\rho^{(n)}(\dots)$, and carry out the usual Taylor series. Now we start by implementing the calculations for the Poisson case, to see the mechanics.

The next term in the expansion is then

$$
\begin{aligned}
V_2(w_1,w_2) &= \mathbf{E}\left\{\sum_v h_p^2(v)\sum_{x,y}^{\neq} h_l(x)h_l(y)\phi_{w_2}(x-y) + \sum_x h_l^2(x)\sum_{v,y}^{\neq} h_p(x)h_p(y)h_l(v)h_l(y)\phi_{w_1}(x-y)\right\}\\
&= \int_{B^3}\rho^{(3)}(x,y,v)h_p^2(v)h_l(x)h_l(y)\phi_{w_2}(x-y)\,dxdydv + \iint_{B^2}\rho^{(2)}(x,y)h_l(x)h_l(y)h_p^2(x)\phi_{w_2}(x-y)\,dxdy\\
&+ \iint_{B^2}\rho^{(2)}(x,y)h_p^2(y)h_l(x)h_l(y)\phi_{w_2}(x-y)\,dxdy + \int_{B^3}\rho^{(3)}(x,y,v)h_l^2(v)h_p(x)h_p(y)\phi_{w_1}(x-y)\,dxdydv\\
&+ \iint_{B^2}\rho^{(2)}(x,y)h_l^2(x)h_p(x)h_p(y)\phi_{w_1}(x-y)\,dxdy + \iint_{B^2}\rho^{(2)}(x,y)h_l^2(y)h_p(x)h_p(y)\phi_{w_1}(x-y)\,dxdy.
\end{aligned}
$$

The final term is in turn using (9)

$$
\begin{aligned}
V_3 &= \mathbf{E}\left[\sum_x h_p^2(x)\sum_y h_l^2(y)\right] = \mathbf{E}\sum_{x\neq y} h_p^2(x)h_l^2(y) + \mathbf{E}\sum_x h_p^2(x)h_l^2(x)\\
&= \int_{B^2}\rho^{(2)}(x,y)h_p^2(x)h_l^2(y)\,dx\,dy + \lambda\int_B h_p^2(x)h_l^2(x)\,dx. \qquad (108)
\end{aligned}
$$

We can put these terms together to determine the covariance between the periodogram at two different wave

numbers and using two different tapers as follows:

$$
\begin{aligned}
\mathbf{Cov}\{I_p^t(w_1), I_l^t(w_2)\} &= \int_{B^4} \rho^{(4)}(x,y,z,v) h_p(x) h_p(y) h_l(z) h_l(v) f_{w_1}(x-y) f_{w_2}(z-v) dx dy dv dz \\
&+ \int_{B^3} \rho^{(3)}(x,y,v) h_p(x) h_p(y) h_l(x) h_l(v) f_{w_1}(x-y) f_{w_2}(x-v) dx dy dv \\
&+ \int_{B^3} \rho^{(3)}(x,y,z) h_p(x) h_p(y) h_l(z) h_l(x) f_{w_1}(x-y) f_{w_2}(z-x) dx dy dz \\
&+ \int_{B^3} \rho^{(3)}(x,y,v) h_p(x) h_p(y) h_l(y) h_l(v) f_{w_1}(x-y) f_{w_2}(y-v) dx dy dv \\
&+ \int_{B^3} \rho^{(3)}(x,y,z) h_p(x) h_p(y) h_l(z) h_l(y) f_{w_1}(x-y) f_{w_2}(z-y) dx dy dz \\
&+ \int_{B^2} \rho^{(2)}(x,y) h_p(x) h_p(y) h_l(x) h_l(y) f_{w_1}(x-y) f_{w_2}(x-y) dx dy \\
&+ \int_{B^2} \rho^{(2)}(x,y) h_p(x) h_p(y) h_l(y) h_l(x) f_{w_1}(x-y) f_{w_2}(y-x) dx dy \\
&+ \int_{B^3} \rho^{(3)}(x,y,v) h_p^2(v) h_l(x) h_l(y) f_{w_2}(x-y) dx dy dv + \iint_{B^2} \rho^{(2)}(x,y) h_p^2(x) h_l(x) h_l(y) f_{w_2}(x-y) dx dy \\
&+ \iint_{B^2} \rho^{(2)}(x,y) h_p^2(y) h_l(x) h_l(y) f_{w_2}(x-y) dx dy + \int_{B^3} \rho^{(3)}(x,y,v) h_l^2(v) h_p(x) h_p(y) f_{w_1}(x-y) dx dy \\
&+ \iint_{B^2} \rho^{(2)}(x,y) h_l^2(x) h_p(x) h_p(y) f_{w_1}(x-y) dx dy + \iint_{B^2} \rho^{(2)}(x,y) h_l^2(y) h_p(x) h_p(y) f_{w_1}(x-y) dx dy \\
&+ \int_{B^2} \rho^{(2)}(x,y) h_p^2(x) h_l^2(y)\, dx\, dy + \lambda \int_B h_p^2(x) h_l^2(x)\, dx \\
&- \left\{ \lambda + \int_{B^2} h_p(x) h_p(y) e^{-i {w_1}^T (x-y)} \rho^{(2)}(x-y) dx dy \right\} \left\{ \lambda + \int_{B^2} h_l(x) h_l(y) e^{-i {w_2}^T (x-y)} \rho^{(2)}(x-y) dx dy \right\} \\
&= \int_{B^4} \rho^{(4)}(x,y,z,v) m^{(4)}(x,y,z,v)\, dx dy dz dv + \int_{B^3} \rho^{(3)}(x,y,z) m^{(3)}(x,y,z)\, dx dy dz \\
&+ \int_{B^2} \rho^{(2)}(x,y) m^{(2)}(x,y)\, dx dy \\
&- \left\{ \lambda + \int_{B^2} h_p(x) h_p(y) e^{-i {w_1}^T (x-y)} \rho^{(2)}(x-y) dx dy \right\} \left\{ \lambda + \int_{B^2} h_l(x) h_l(y) e^{-i {w_2}^T (x-y)} \rho^{(2)}(x-y) dx dy \right\},
\end{aligned}
$$

this defining $m^{(4)}(x,y,z,v)$, $m^{(3)}(x,y,z)$ and $m^{(2)}(x,y)$, respectively.

We start by noting that the multiplier of $\rho^{(3)}$ takes the form of

$$
\begin{aligned}
m^{(3)}(x,y,z) &= h_p(x) h_p(y) h_l(x) h_l(z) \phi_{w_1}(x-y) \phi_{w_2}(x-z) + h_p(x) h_p(y) h_l(z) h_l(x) \phi_{w_1}(x-y) \phi_{-w_2}(x-z) \\
&+ h_p(x) h_p(y) h_l(y) h_l(z) \phi_{w_1}(x-y) \phi_{w_2}(y-z) + h_p(x) h_p(y) h_l(z) h_l(y) \phi_{w_1}(x-y) \phi_{-w_2}(y-z) \\
&+ h_p^2(z) h_l(x) h_l(y) \phi_{w_2}(x-y) + h_l^2(z) h_p(x) h_p(y) \phi_{w_1}(x-y) \\
&= h_p(x) h_p(y) h_l(x) h_l(z) \phi_{w_1}(x-y) \{\phi_{w_2}(x-z) + \phi_{-w_2}(x-z)\} \\
&+ h_p(x) h_p(y) h_l(y) h_l(z) \phi_{w_1}(x-y) \{\phi_{w_2}(y-z) + \phi_{-w_2}(y-z)\} \\
&+ \phi_{w_2}(x-y) h_p^2(z) h_l(x) h_l(y) + \phi_{w_1}(x-y) h_l^2(z) h_p(x) h_p(y)
\end{aligned}
$$

Looking at the positive term multiplying $\rho^{(2)}(x,y)$ we can simplify to

$$
\begin{aligned}
m^{(2)}(x,y) &= h_p(x) h_p(y) h_l(x) h_l(y) \phi_{w_1}(x-y) \phi_{w_2}(x-y) + h_p(x) h_p(y) h_l(y) h_l(x) \phi_{w_1}(x-y) \phi_{w_2}(y-x) \\
&+ h_p^2(x) h_l(x) h_l(y) \phi_{w_2}(x-y) + h_p^2(y) h_l(x) h_l(y) \phi_{w_1}(x-y) + h_l^2(x) h_p(x) h_p(y) \phi_{w_1}(x-y) \\
&+ h_l^2(x) h_p(x) h_p(y) \phi_{w_2}(x-y) + h_p^2(x) h_l^2(y) \\
&= h_p(x) h_p(y) h_l(x) h_l(y) f_{w_1}(x-y) \{f_{w_2}(x-y) + f_{-w_2}(x-y)\} + h_p^2(x) h_l(x) h_l(y) \\
&\times \{f_{w_2}(x-y) + f_{w_1}(x-y)\} + h_l^2(x) h_p(x) h_p(y) \{f_{w_2}(x-y) + f_{w_1}(x-y)\} + h_p^2(x) h_l^2(y) \\
&= h_p(x) h_p(y) h_l(x) h_l(y) \phi_{w_1}(x-y) \{\phi_{w_2}(x-y) + \phi_{-k_2}(x-y)\} + \left\{h_p^2(x) h_l(x) h_l(y) + h_l^2(x) h_p(x) h_p(y)\right\} \\
&\cdot \{\phi_{w_2}(x-y) + \phi_{w_1}(x-y)\} + h_p^2(x) h_l^2(y).
\end{aligned}
$$

Then it follows that we can simplify the final term:

$$\left\{\lambda + \int_{B^2} h_p(x)h_p(y)e^{-iw_1{}^T(x-y)}\rho^{(2)}(x-y)dxdy\right\}\left\{\lambda + \int_{B^2} h_l(x)h_l(y)e^{-iw_2{}^T(x-y)}\rho^{(2)}(x-y)dxdy\right\}$$
$$= \lambda^2 + \lambda\int_{B^2}(h_p(x)h_p(y)\phi_{w_1}(x-y) + h_l(x)h_l(y)\phi_{w_2}(x-y))\rho^{(2)}(x-y)dxdy$$
$$+ \left\{\int_{B^2} h_p(x)h_p(y)\phi_{w_1}(x-y)\rho^{(2)}(x-y)dxdy\right\}\left\{\int_{B^2} h_l(x)h_l(y)\phi_{w_2}(x-y)\rho^{(2)}(x-y)dxdy\right\}$$

We then get for the tapered periodogram:

$$\begin{aligned}
\mathbf{Cov}\{I_p^t(w_1), I_l^t(w_2)\} &= \int_{B^4}\rho^{(4)}(x,y,z,v)h_p(x)h_p(y)h_l(z)h_l(v)\phi_{w_1}(x-y)\phi_{w_2}(z-v)\,dxdydvdz \\
&+ \int_{B^3}\rho^{(3)}(x,y,v)\left[h_p(x)h_p(y)h_l(x)h_l(v)\phi_{w_1}(x-y)\left\{\phi_{w_2}(x-v)+\phi_{-w_2}(x-v)\right\}\right. \\
&+ h_p(x)h_p(y)h_l(y)h_l(v)\phi_{w_1}(x-y)\left\{\phi_{w_2}(y-v)+\phi_{-w_2}(y-v)\right\} \\
&\left.+\phi_{w_2}(x-y)h_p^2(v)h_l(x)h_l(y) + \phi_{w_1}(x-y)h_l^2(v)h_p(x)h_p(y)\right]\,dxdydv \\
&+ \int_{B^2}\rho^{(2)}(x,y)\left[h_p(x)h_p(y)h_l(x)h_l(y)\phi_{w_1}(x-y)\left\{\phi_{w_2}(x-y)+\phi_{-w_2}(x-y)\right\}\right. \\
&\left.+\left\{h_p^2(x)h_l(x)h_l(y)+h_l^2(x)h_p(x)h_p(y)\right\}\left\{\phi_{w_2}(x-y)+\phi_{w_1}(x-y)\right\}+h_p^2(x)h_l^2(y)\right]\,dxdy \\
&+ \lambda\int_B h_p^2(x)h_l^2(x)\,dx - \lambda^2 - \lambda\int_{B^2}h_p(x)h_p(y)h_l(x)h_l(y)(\phi_{w_1}(x-y)+\phi_{w_2}(x-y))\rho^{(2)}(x,y)dxdy \\
&- \left\{\int_{B^2}h_p(x)h_p(y)\phi_{w_1}(x-y)\rho^{(2)}(x,y)dxdy\right\}\cdot\left\{\int_{B^2}h_l(x)h_l(y)\phi_{w_2}(x-y)\rho^{(2)}(x,y)dxdy\right\}.
\end{aligned}$$

This gives the covariance between the tapered periodogram with different tapers, at different wave numbers, and it quite a useful general expression, as the following expression will show. □

I Covariance Terms for the Bias–Corrected Periodogram

Proof. This is a job for Campbell's theorem, in several instances. We need to compute (i) $\mathbf{Cov}\{I_h(w_1), I_h(w_2)\}$, (ii) $\mathbf{Cov}\{I_h(w_1), \Re\{J_h(w_2)H^*(w_2)\}\}$, (iii) $\mathbf{Cov}\{I_h(w_2), \Re\{J_h(w_1)H^*(w_1)\}\}$ and (iv) $\mathbf{Cov}\{\Re\{J_h(w_1)H^*(w_1)\}, \Re\{J_h(w_2)H^*(w_2)\}\}$.

Before we calculate this exactly, we note that from (14) as $J_h(w) \to N_C(\lambda H(w), f(w), r(w))$, with

$$r(w) = \lambda\int_{\mathbb{R}^d} H(k'-2w)H(k')dk' + \int_{\mathbb{R}^d} U(w,z)e^{-2\pi iw\cdot z}\{\rho(z)-\lambda^2\}dxdz.$$

We can see in increasing w that $|r(w)| \to 0$. Assuming that $J_h(w)$ is uniformly integrable, then we can use the limiting behaviour coupled with Isserliss' (Wick's) theorem (Schreier and Scharf, 2003, 2010) to deduce that when $\lambda H(w) \approx 0$ (Isserliss' theorem applied for zero-mean Gaussian random variables):

$$\begin{aligned}
\mathbf{Cov}\{I_h(w_1), I_h(w_2)\} &= \mathbf{E}\{J_h(w_1)J_h^*(w_1)J_h(w_2)J_h^*(w_2)\} - \mathbf{E}\{J_h(w_1)J_h^*(w_1)\}\mathbf{E}\{J_h(w_2)J_h^*(w_2)\} \\
&= \mathbf{E}\{J_h(w_1)J_h^*(w_1)\}\mathbf{E}\{J_h(w_2)J_h^*(w_2)\} + \mathbf{E}\{J_h(w_1)J_h(w_2)\}\mathbf{E}\{J_h^*(w_1)J_h^*(w_2)\} \\
&+ \mathbf{E}\{J_h(w_1)J_h^*(w_2)\}\mathbf{E}\{J_h^*(w_1)J_h(w_2)\} - \mathbf{E}\{J_h(w_1)J_h^*(w_1)\}\mathbf{E}\{J_h(w_2)J_h^*(w_2)\} \\
&= \mathbf{E}\{J_h(w_1)J_h(w_2)\}\mathbf{E}\{J_h^*(w_1)J_h^*(w_2)\} + \mathbf{E}\{J_h(w_1)J_h^*(w_2)\}\mathbf{E}\{J_h^*(w_1)J_h(w_2)\}. \qquad (109)
\end{aligned}$$

A problem is now that whilst we have marginal convergence at a single fixed w, we have not specified the joint convergence. We can apply Campbell's theorem and will need to recall that $\lambda H(w) \approx 0$ when comparing it to the expression in Theorem 5.1. In fact, if we can assume that $\lambda H(w) \approx 0$, the bilinear operation simplifies. This will be true at higher wave numbers, but that makes the argument circular. □

J Proof of Proposition 5.1

Proof. This proposition both determines the variance of a single taper periodogram, and the cross-covariance between the tapers at a fixed wavenumber w for a Poisson Process. If we start from the assumption of Poissonian

(namely (44)) then we find

$$\begin{aligned}
\mathbf{Cov}\{I_p^t(w), I_l^t(w)\} &= \int_{B^4} \lambda^4 h_p(x)h_p(y)h_l(z)h_l(v)\phi_w(x-y)\phi_w(z-v)\,dxdydvdz \\
&+ \int_{B^3} \lambda^3 \left[h_p(x)h_p(y)h_l(x)h_l(v)\phi_w(x-y)\left\{\phi_w(x-v)+\phi_{-w}(x-v)\right\}\right. \\
&+ h_p(x)h_p(y)h_l(y)h_l(v)\phi_w(x-y)\left\{\phi_w(y-v)+\phi_{-w}(y-v)\right\} \\
&\left.+\phi_w(x-y)h_p^2(v)h_l(x)h_l(y)+\phi_w(x-y)h_l^2(v)h_p(x)h_p(y)\right]\,dxdydv \\
&+ \int_{B^2} \lambda^2 \left[h_p(x)h_p(y)h_l(x)h_l(y)\phi_w(x-y)\left\{\phi_w(x-y)+\phi_{-w}(x-y)\right\}\right. \\
&\left.+\left\{h_p^2(x)h_l(x)h_l(y)+h_l^2(x)h_p(x)h_p(y)\right\}\left\{\phi_w(x-y)+\phi_w(x-y)\right\}+h_p^2(x)h_l^2(y)\right]\,dxdy \\
&+ \lambda\int_B h_p^2(x)h_l^2(x)\,dx - \lambda^2 - \lambda\int_{B^2} h_p(x)h_p(y)h_l(x)h_l(y)(\phi_w(x-y)+\phi_w(x-y))\lambda^2 dxdy \\
&- \left\{\int_{B^2} h_p(x)h_p(y)\phi_k(x-y)\lambda^2 dxdy\right\}\cdot\left\{\int_{B^2} h_l(x)h_l(y)\phi_w(x-y)\lambda^2 dxdy\right\}.
\end{aligned}$$

We now need to massage this expression in order to understand the correlation. We first note that the last and first terms cancel exactly, leaving us with

$$\begin{aligned}
\mathbf{Cov}\{I_p^t(w), I_l^t(w)\} &= \lambda^3\int_{B^3} \left[h_p(x)h_p(y)h_l(x)h_l(v)\phi_w(x-y)\left\{\phi_w(x-v)+\phi_{-w}(x-v)\right\}\right. \\
&\left.+h_p(x)h_p(y)h_l(y)h_l(v)\phi_w(x-y)\left\{\phi_w(y-v)+\phi_{-w}(y-v)\right\}\right]\,dxdydv \\
&+ \lambda^3\int_{B^2} \phi_w(x-y)h_l(x)h_l(y)+\phi_w(x-y)h_p(x)h_p(y)\,dxdy \\
&+ \lambda^2\int_{B^2} \left[h_p(x)h_p(y)h_l(x)h_l(y)\phi_w(x-y)\left\{\phi_w(x-y)+\phi_{-w}(x-y)\right\}\right. \\
&\left.+\left\{h_p^2(x)h_l(x)h_l(y)+h_l^2(x)h_p(x)h_p(y)\right\}2\phi_w(x-y)\right]\,dxdy \\
&+ \lambda\int_B h_p^2(x)h_l^2(x)\,dx - 2\lambda^3\int_{B^2} h_p(x)h_p(y)h_l(x)h_l(y)\phi_w(x-y)\,dxdy \\
&= \sum_{j=1}^7 T_j(w),
\end{aligned}$$

this defining the sequence $\{T_j(w)\}$. We need some extra relationships to simplify these expressions. We note that

$$\int_B h_p(x)h_l(x)\,dx = \delta_{pl} \tag{110}$$

$$\int_B h_p(x)\phi_{2w}(x)\,dx \approx 0, \quad w > b_w > 0, \tag{111}$$

$$\int H_p(w'-2w)H_l(w')\,dw' \approx 0 \quad w > b_w > 0, \tag{112}$$

where b_w is the bandwidth of the window. We find

$$\begin{aligned}
T_1(w) &= \lambda^3\int_{B^3} h_p(x)h_p(y)h_l(x)h_l(v)\phi_w(x-y)\left\{\phi_w(x-v)+\phi_{-w}(x-v)\right\}\,dxdydv \\
&= \lambda^3 H_p(-w)H_l(-w)\int H_p(w'-2w)H_l(w')\,dw' + \lambda^3\delta_{pl}H_p(w)H_l(-w).
\end{aligned} \tag{113}$$

The next term is

$$\begin{aligned}
T_2(w) &= \lambda^3\int_{B^3} h_p(x)h_p(y)h_l(y)h_l(v)\phi_w(x-y)\left\{\phi_w(y-v)+\phi_{-w}(y-v)\right\}\,dxdydv \\
&= \lambda^3\delta_{pl}H_p(w)H_l(-w) + \lambda^3 H_p(-w)H_l(-w)\int H_p(w'-2w)H_l(w')\,dw'.
\end{aligned} \tag{114}$$

The following term is

$$\begin{aligned}
T_3(w) &= \lambda^3\int_{B^2} \phi_w(x-y)h_l(x)h_l(y)+\phi_w(x-y)h_p(x)h_p(y)\,dxdy \\
&= \lambda^3 H_l(w)H_l(-w) + \lambda^3 H_p(w)H_p(-w).
\end{aligned}$$

The next term is

$$T_4(w) = \lambda^2 \int_{B^2} [h_p(x)h_p(y)h_l(x)h_l(y)\phi_w(x-y)\{\phi_w(x-y)+\phi_{-w}(x-y)\}]\,dxdy$$
$$= \lambda^2 \left| \int H_p(w'-2w)H_l(w')\,dw' \right|^2 + \lambda^2\delta_{pl}.$$

The next term then takes the form

$$T_5(w) = 2\lambda^2 \int_{B^2} \left\{h_p^2(x)h_l(x)h_l(y) + h_l^2(x)h_p(x)h_p(y)\right\}\phi_w(x-y)\,dxdy$$
$$= 2\lambda^2 \left\{ H_l(-w)\int_{\mathbb{R}^{2d}} H_p(w'')H_p(w'-w'')H_l(w-w')\,dw'dw'' + H_p(-w)\int_{\mathbb{R}^{2d}} H_l(w'')H_l(w'-w'')H_p(w-w')\,dw'dw'' \right\}.$$

The next term takes the form of

$$T_6(w) = \lambda \int_B h_p^2(x)h_l^2(x)\,dx \sim \frac{\lambda}{|B|}.$$

And the final term is negative:

$$T_7(w) = -2\lambda^3 \int_{B^2} h_p(x)h_p(y)h_l(x)h_l(y)\phi_w(x-y)\,dxdy$$
$$= -2\lambda^3 \left| \int H_p(w-w')H_l(w')\,dw' \right|^2.$$

As we have assumed the wave number w is sufficiently large both $|H_p(w)| \approx |H_l(w)| \approx 0$. This means the only surviving contributions are

$$\mathbf{Cov}\{I_p^t(w), I_l^t(w)\} \approx \lambda \int_B h_p^2(x)h_l^2(x)\,dx + \lambda^2\delta_{pl}.$$

This yields the stated result. We can derive the $o(1)$ terms formally should we wish by quantifying the leakage in ℓ, the length of a side. □

K Proof of Proposition 5.2

Proof. If we start from the assumption of Poissonian (namely (44)) then

$$\begin{aligned}
\mathbf{Cov}\{I_p^t(w_1), I_p^t(w_2)\} &= \int_{B^4} \lambda^4 h_p(x)h_p(y)h_p(z)h_p(v)\phi_{w_1}(x-y)\phi_{w_2}(z-v)\,dxdydvdz \\
&+ \int_{B^3} \lambda^3 \left[h_p^2(x)h_p(y)h_p(v)\phi_{w_1}(x-y)\{\phi_{w_2}(x-v)+\phi_{-w_2}(x-v)\}\right. \\
&+ h_p(x)h_p^2(y)h_p(v)\phi_{w_1}(x-y)\{\phi_{w_2}(y-v)+\phi_{-w_2}(y-v)\} \\
&\left.+\phi_{w_2}(x-y)h_p^2(v)h_p(x)h_p(y) + \phi_{w_1}(x-y)h_p^2(v)h_p(x)h_p(y)\right]\,dxdydv \\
&+ \int_{B^2} \lambda^2 \left[h_p(x)h_p(y)h_p(x)h_p(y)\phi_{w_1}(x-y)\{\phi_{w_2}(x-y)+\phi_{-w_2}(x-y)\}\right. \\
&\left.+ \left\{h_p^2(x)h_p(x)h_p(y) + h_p^2(x)h_p(x)h_p(y)\right\}\{\phi_{w_2}(x-y)+\phi_{w_1}(x-y)\} + h_p^2(x)h_p^2(y)\right]\,dxdy \\
&+ \lambda\int_B h_p^4(x)\,dx - \lambda^2 - \lambda\int_{B^2} h_p^2(x)h_p^2(y)(\phi_{w_1}(x-y)+\phi_{w_2}(x-y))\lambda^2 dxdy \\
&- \left\{\int_{B^2} h_p(x)h_p(y)\phi_{w_1}(x-y)\lambda^2 dxdy\right\}\cdot\left\{\int_{B^2} h_p(x)h_p(y)\phi_{w_2}(x-y)\lambda^2 dxdy\right\}
\end{aligned}$$

Continuing on with the calculations we find:

$$\begin{aligned}
\mathbf{Cov}\{I_p^t(w_1), I_p^t(w_2)\} &= \int_{B^3} \lambda^3 \left[h_p^2(x)h_p(y)h_p(v)\phi_{w_1}(x-y)\{\phi_{w_2}(x-v)+\phi_{-w_2}(x-v)\}\right. \\
&+ h_p(x)h_p^2(y)h_p(v)\phi_{w_1}(x-y)\{\phi_{w_2}(y-v)+\phi_{-w_2}(y-v)\} \\
&\left.+\phi_{w_2}(x-y)h_p^2(v)h_p(x)h_p(y) + \phi_{w_1}(x-y)h_p^2(v)h_p(x)h_p(y)\right]\,dxdydv \\
&+ \int_{B^2} \lambda^2 \left[h_p^2(x)h_p^2(y)\phi_{w_1}(x-y)\{\phi_{w_2}(x-y)+\phi_{-w_2}(x-y)\}\right. \\
&\left.+ \left\{h_p^3(x)h_p(y) + h_p^3(x)h_p(y)\right\}\{\phi_{w_2}(x-y)+\phi_{w_1}(x-y)\}\right]\,dxdy \\
&+ \lambda\int_B h_p^4(x)\,dx - \lambda^3\int_{B^2} h_p^2(x)h_p^2(y)(\phi_{w_1}(x-y)+\phi_{w_2}(x-y))dxdy.
\end{aligned}$$

We like in the previous case split this into seven parts $\{\widetilde{T}_j(w)\}$. We find that

$$\begin{aligned}\widetilde{T}_1(w) &= \lambda^3 \int_{B^3} h_p^2(x) h_p(y) h_p(v) \phi_{w_1}(x-y) \left\{\phi_{w_2}(x-v) + \phi_{-w_2}(x-v)\right\} \, dxdydv \\ &= \lambda^3 H_p(-w_1) H_p(-w_2) \int_{\mathbb{R}^d} H_p(w') H_p(w_1 + w_2 - w') \, dw' \\ &+ \lambda^3 H_p(-w_1) H_p(w_2) \int_{\mathbb{R}^d} H_p(w') H_p(w_1 - w_2 - w') \, dw'.\end{aligned}$$

If $w_1 - w_2, w_1 + w_2 \neq 0$ and is some fixed number, then this becomes negligible. The next term is

$$\begin{aligned}\widetilde{T}_2(w) &= \lambda^3 \int_{B^3} h_p(x) h_p^2(y) h_p(v) \phi_{w_1}(x-y) \left\{\phi_{w_2}(y-v) + \phi_{-w_2}(y-v)\right\} dxdydv \\ &= \lambda^3 H_p(w_1) H_p(-w_2) \int_{\mathbb{R}^d} H_p(w') H_p(w_2 - w_1 - w') \, dw' \\ &+ \lambda^3 H_p(w_1) H_p(-w_2) \int_{\mathbb{R}^d} H_p(w') H_p(w_2 + w_1 - w') \, dw'.\end{aligned}$$

If $w_1 - w_2, w_1 + w_2 \neq 0$ and is some fixed number, then this becomes negligible. The next term is

$$\begin{aligned}\widetilde{T}_3(w) &= \lambda^3 \int_{B^3} \left[\phi_{w_2}(x-y) h_p^2(v) h_p(x) h_p(y) + \phi_{w_1}(x-y) h_p^2(v) h_p(x) h_p(y)\right] \, dxdydv \\ &= \lambda^3 \left|H_p(w_2)\right|^2 + \lambda^3 \left|H_p(w_1)\right|^2.\end{aligned}$$

As $w_1, w_2 > b_w$ the bandwidth this contribution becomes negligible. The next contribution in the expression takes the form

$$\begin{aligned}\widetilde{T}_4(w) &= \int_{B^2} \lambda^2 \left[h_p^2(x) h_p^2(y) \phi_{w_1}(x-y) \left\{\phi_{w_2}(x-y) + \phi_{-w_2}(x-y)\right\}\right] dxdy \\ &= \lambda^2 \left| \int_{\mathbb{R}^d} H_p(w') H_p(w_1 + w_2 - w') \, dw' \right|^2 + \lambda^2 \left| \int_{\mathbb{R}^d} H_p(w') H_p(w_1 - w_2 - w') \, dw' \right|^2.\end{aligned}$$

Then the next term is

$$\widetilde{T}_5(w) = 2\lambda^2 \int_{B^2} h_p^3(x) h_p(y) \, \left\{\phi_{w_2}(x-y) + \phi_{w_1}(x-y)\right\}] \, dxdy.$$

Figuring out the size of this contribution is a little bit more complex. The integral will do a Fourier transform in y of $h_p(y)$ in w_1 and w_2, and also the conjugate will be taken. The Fourier transform in y will be supported when $|w_1| < b_w$ and $|w_2| < b_w$, but will not be supported when either magnitude gets too large. The second Fourier transform in x will be somewhat more spread–this is because we are Fourier transforming $h_p^3(x)$, which will be a third order convolution once Fourier transformed. But as long as we can use the joint concentration in $|w_1| < b_w$ and $|k_2| < b_k$ this will still be negligible.

The next term is given by

$$\widetilde{T}_6(w) = \lambda \int_B h_p^4(x) \, dx = \lambda \|h_p\|_4^4 = o(1).$$

To see why this is true, note that if $h_p(x)$ is constant then we can easily calculate the higher order norms. We find by using the Cauchy–Schwarz inequality that if $p \neq q$ that

$$\begin{aligned}T_6^2(w) &= \left[\int_B h_p^2(x) h_l^2(x) \, dx\right]^2 \\ &\leq \int_B h_p^4(x) \int_B h_l^4(x) \\ \Rightarrow 0 \leq \int_B h_p^2(x) h_l^2(x) \, dx &\leq \max\{\int_B h_p^4(x), \int_B h_l^4(x)\}. \qquad (115)\end{aligned}$$

We also note that

$$1^2 = \left[\int_B 1 \cdot h^2(x) \, dx\right]^2 \leq \int_B 1^2 \, dx \cdot \int_B h^4(x) \, dx \qquad (116)$$

$$\Rightarrow \frac{1}{|B|} \leq \int_B h^4(x) \, dx. \qquad (117)$$

We now need to determine an upper bound. We then look at

$$\left[\int_B h^4(x)\,dx\right]^2 \leq \int_B 1^2\,dx \cdot \int_B h^8(x)\,dx = |B| \cdot \int_B h^8(x)\,dx. \tag{118}$$

Then we have

$$\frac{1}{|B|} \leq \int_B h^4(x)\,dx \leq \sqrt{|B| \cdot \int_B h^8(x)\,dx}. \tag{119}$$

We can for any taper $h(x)$ calculate $\int_B h^4(x)\,dx$ explicitly. For the constant function we get $\frac{1}{|B|}$. For tapers well–concentrated we would expect a similar decrease, but for any choice of taper we can calculate the value of the 4th norm explicitly.

Thus we understand a bit more about this term. Moving on to the next aspect of the computation, finally,

$$\begin{aligned}\widetilde{T}_7(w) &= -\lambda^3 \int_{B^2} h_p^2(x)h_p^2(y)(\phi_{w_1}(x-y)+\phi_{w_2}(x-y))dxdy \\ &= -\lambda^3\left|\int_{\mathbb{R}^d} H_p(w')H_p(w_1-w')\,dw'\right|^2 - \lambda^3\left|\int_{\mathbb{R}^d} H_p(w')H_p(w_2-w')\,dw'\right|^2.\end{aligned}$$

This shows each individual contribution as long as $|w_1-w_2|, |w_1+w_2| > b_w$. This completes the proof. □

L Proof of Theorem 5.2

We start from Theorem 5.1. We use that we have set $\phi_w(x) = e^{-2\pi i w\cdot x}$. We then have:

$$\begin{aligned}\mathbf{Var}\{I_h(w)\} = &\overbrace{\int_{B^4} \rho^{(4)}(x,y,z,v)h(x)h(y)h(z)h(v)\phi_w(x-y)\phi_w(z-v)\,dxdydvdz}^{T_0(w)} \\ &+\overbrace{\int_{B^3} \rho^{(3)}(x,y,v)\left[h(x)h(y)h(x)h(v)\phi_w(x-y)\left\{\phi_w(x-v)+\phi_{-w}(x-v)\right\}\right.}^{T_1(w)} \\ &+\overbrace{h(x)h(y)h(y)h(v)\phi_w(x-y)\left\{\phi_w(y-v)+\phi_{-w}(y-v)\right\}}^{T_2(w)} \\ &\left.+\overbrace{\phi_w(x-y)h^2(v)h(x)h(y)+\phi_w(x-y)h^2(v)h(x)h(y)}^{T_3(w)}\right]dxdydv \\ &+\overbrace{\int_{B^2} \rho^{(2)}(x,y)\left[h(x)h(y)h(x)h(y)\phi_w(x-y)\left\{\phi_w(x-y)+\phi_{-w}(x-y)\right\}\right.}^{T_4(w)} \\ &+\overbrace{\left\{h^2(x)h(x)h(y)+h^2(x)h(x)h(y)\right\}\left\{\phi_w(x-y)+\phi_w(x-y)\right\}+h^2(x)h^2(y)\left.\right]\,dxdy}^{T_5(w)} \\ &+\lambda\overbrace{\int_B h^2(x)h^2(x)\,dx - \lambda^2 - \lambda\int_{B^2} h(x)h(y)h(x)h(y)(\phi_w(x-y)+\phi_w(x-y))\rho^{(2)}(x,y)dxdy}^{T_6(w)} \\ &-\overbrace{\left\{\int_{B^2} h(x)h(y)\phi_w(x-y)\rho^{(2)}(x,y)dxdy\right\}\cdot\left\{\int_{B^2} h(x)h(y)\phi_w(x-y)\rho^{(2)}(x,y)dxdy\right\}}^{T_7(w)}.\end{aligned} \tag{120}$$

We wish to simplify this expression to get a sense of the major contribution to the variance. We recall that we are analyzing a stationary process and so with some abuse of notation

$$\rho^{(n)}(x_2-x_1,\ldots,x_n-x_1) = \rho^{(n)}(x_1,x_2,\ldots,x_n). \tag{121}$$

First we define

$$\bar{\rho}^{(n)}(\tau_1,\ldots,\tau_{n-1}) = \frac{\bar{\rho}^{(n)}(\tau_1,\ldots,\tau_{n-1})-\lambda^n}{\lambda^n}. \tag{122}$$

For a Poisson process these functions are all identically zero. We define the Fourier transform of these functions $\{\widetilde{\rho}^{(n)}(\tau_1,\ldots,\tau_{n-1})\}$ to be

$$\widetilde{f}^{(n)}(\boldsymbol{w}) = \int_{\mathbb{R}^{(n-1)d}} e^{-2\pi i \tau^T \boldsymbol{w}} \widetilde{\rho}^{(n)}(\tau)\, d\tau. \tag{123}$$

These are higher order generalizations of the common spectrum. Apart from possibly at $\boldsymbol{w}=0$ we assume that these functions are twice differentiable, just like we did for the centred spectrum. We can now exchange the representation of $\mathbf{Var}\{I_h(w)\}$ using $\widetilde{f}^{(n)}(\boldsymbol{w})$, rather like our treatment of $\mathbf{E}\{I_h(w)\}$ in Theorem 4.1; and then as a last step correct with the effect of a Poisson process.

Referring back to (120) we see that all terms are multiplications of $\rho^{(n)}(\cdot)$ with products of $h(\cdot)$ in space. These in the Fourier domain become convolutions between the spectrum and higher order generalizations and the window functions (products of $h(x)$). This follows from the convolution theorem.

To arrive at a simplified expression, we first transform between $\bar{\rho}(\tau)$ and $\widetilde{\rho}(\tau)$. This adds exactly the variance term of a Poisson process, in all places but the final line of (120) which is the only place $\rho(\tau)$ does not appear linearly. Once the representation is in terms of $\widetilde{\rho}(\tau)$, then we transfer the product to a convolution in the Fourier domain.

The term involving $\widetilde{f}^{(n)}(w)$ is assumed smooth (apart from at $w=0$), and as the convolution involves the windows $h(x)$, we assume the convolution is local. A Fourier transform of (120) term by term, using the notation of the variance calculation for a Poisson process yields with added terms $T_0(w)$ and $T_8(w)$. Let us show each term

$$\begin{aligned}
T_0(w) &= \int_{B^4} \rho^{(4)}(x,y,z,v)h(x)h(y)h(z)h(v)\phi_w(x-y)\phi_w(z-v)\, dxdydvdz \\
&= \int \left(\rho^{(4)}(x,y,z,v) - \lambda^4 + \lambda^4\right) h(x)h(y)h(z)h(v)\phi_w(x-y)\phi_w(z-v)\, dxdydvdz \\
&= \int_{B^4} \rho^{(4)}(x,y,z,v)h(x)h(y)h(z)h(v)\phi_w(x-y)\phi_w(z-v)\, dxdydvdz \\
&= \lambda^4 \int \left\{1 + \widetilde{\rho}^{(4)}(x-v,y-v,z-v)\right\} h(x)h(y)h(z)h(v)\phi_w(x-v+v-y)\phi_w(z-v)\, dxdydvdz.
\end{aligned}$$

We can now illustrate how the integrals behave. We start with the change of variables

$$\tau_1 = x-v, \quad \tau_2 = y-v, \quad \tau_3 = z-v. \tag{124}$$

Then we find

$$T_0(w) = \lambda^4 \int \left\{1 + \widetilde{\rho}^{(4)}(\tau_1,\tau_2,\tau_3)\right\} h(\tau_1+v)h(\tau_2+v)h(\tau_3+v)h(v)\phi_w(\tau_1-\tau_2)\phi_w(\tau_3)\, d\tau_1 d\tau_2 d\tau_3 dv.$$

We now need to use some of the large sample characteristics.

$$\begin{aligned}
T_0(w) &= \lambda^4 \widetilde{f}^{(4)}(\boldsymbol{w}) + \lambda^4 \\
T_1(w) &= o(1) \\
T_2(w) &= o(1) \\
T_3(w) &= o(1) \\
T_4(w) &= \lambda^2 \widetilde{f}^{(2)}(w) + \lambda^2 \\
T_5(w) &= o(1) \\
T_6(w) &= \lambda \|h\|_4^4 = o(1) \\
T_7(w) &= o(1) \\
T_8(w) &= \lambda^4 \left\{\widetilde{f}^{(2)}(w) + 1\right\}^2.
\end{aligned} \tag{125}$$

We are left with terms T_4 and T_6 as well as T_0 and T_8.

$$\begin{aligned}
&\mathbf{Var}\{I_h(w)\} \sim \lambda^4 \widetilde{f}^{(4)}(w) + \lambda^4 + \lambda^2 \widetilde{f}^{(2)}(w) + \lambda^2 + \lambda\|h\|_4^4 - \left(\lambda^2 \widetilde{f}^{(2)}(w) + \lambda^2\right)^2 \\
&= \lambda^4 \widetilde{f}^{(4)}(w) + \lambda^4 + \lambda^2 \widetilde{f}^{(2)}(w) + \lambda^2 + \lambda\|h\|_4^4 - \lambda^4 (\widetilde{f}^{(2)}(w))^2 - 2\lambda^4 \widetilde{f}^{(2)}(w) - \lambda^4 \\
&= \lambda^4 \widetilde{f}^{(4)}(w) + \lambda^2 \widetilde{f}^{(2)}(w) + \lambda^2 + \lambda\|h\|_4^4 - \lambda^4 (\widetilde{f}^{(2)}(w))^2 - 2\lambda^4 \widetilde{f}^{(2)}(w) \\
&= \lambda^4 \left\{\widetilde{f}^{(4)}(w) - \widetilde{f}^{(2)}(w)(2 + \widetilde{f}^{(2)}(w))\right\} + \lambda^2 \{\widetilde{f}^{(2)}(w) + 1\} + \lambda\|h\|_4^4.
\end{aligned}$$

Assuming the dominance of the second term in the expression we get

$$\mathbf{Var}\{I_h(w)\} \sim \lambda^2 \widetilde{f}^{(2)}(w) + \lambda^2 = f(w) - \lambda + \lambda^2.$$

We see that the variance increases strictly with the rate $\lambda > 1$.

M 4th Order Moment of Taper

We start by using the Cauchy–Schwarz inequality to obtain

$$\begin{aligned} T_6^2(w) &= \left[\int_B h_p^2(x) h_l^2(x)\, dx\right]^2 \\ &\leq \int_B h_p^4(x) \int_B h_l^4(x) \\ \Rightarrow 0 \leq \int_B h_p^2(x) h_l^2(x)\, dx &\leq \max\{\int_B h_p^4(x), \int_B h_l^4(x)\}. \end{aligned} \tag{126}$$

We also note that

$$\begin{aligned} 1^2 = \left[\int_B 1 \cdot h^2(x)\, dx\right]^2 &\leq \int_B 1^2\, dx \cdot \int_B h^4(x)\, dx, \\ \Rightarrow \frac{1}{|B|} &\leq \int_B h^4(x)\, dx. \end{aligned} \tag{127}$$

We now need to determine an upper bound. We then look at

$$\left[\int_B h^4(x)\, dx\right]^2 \leq \int_B 1^2\, dx \cdot \int_B h^8(x)\, dx = |B| \cdot \int_B h^8(x)\, dx. \tag{128}$$

Then we have

$$\frac{1}{|B|} \leq \int_B h^4(x)\, dx \leq \sqrt{|B| \cdot \int_B h^8(x)\, dx}. \tag{129}$$

We can for any taper $h(x)$ calculate $\int_B h^4(x)\, dx$ explicitly. For the constant function we get $\frac{1}{|B|}$. For tapers well–concentrated we would expect a similar decrease, but for any choice of taper we can calculate the value of the 4th norm explicitly.

N Proof of Lemma 6.1

The estimate in (53) smooths in a local region of w as determined in shape by W, and scale by $\mathbf{\Omega}$. We assume either that either $W(w) = W(w_1) \dots W(w_d)$ and that $W(w_j) = W(-w_j)$ for $1 \leq j \leq d$ (e.g. that the function is separable and even) or that $W(w) = W(|w|)$ (e.g. that the function is isotropic). We can then determine the bias and variance of this estimator using Theorems 4.1 and 5.2, which will be the next step in our study.

As a first step we calculate

$$\mathbf{E}\left\{\overline{I}_h(w; \mathbf{\Omega})\right\} = \sum_{w' \in \mathcal{K}(B_\square(\boldsymbol{\iota}))} \mathbf{E}\left\{\widetilde{I}_h(w')\right\} W_{\mathbf{\Omega}}(w - w'). \tag{130}$$

First we need to settle the normalization of $W(w)$, and make that consistent when going between continuous and discrete operations.

We note that if $w \neq 0$ then we may assume that $\widetilde{f}(w)$ is twice differentiable and so we may appeal to Theorem 4.1 to quantify $\mathbf{E}\left\{\overline{I}_h(w; \mathbf{\Omega})\right\}$. We need to be a little bit more careful with the order term because it will matter.

Referring back to Corollary 4.2 we may note

$$\mathbf{E}\widetilde{I}_h(w) = f(w) + e(w).$$

Here $e(w)$ is defined as the error, but we can deduce its order from the proof of Theorem 4.1.

We therefore write for symmetric kernel function $W(\cdot)$ with diagonal bandwidth matrix $\mathbf{\Omega}$:

$$\begin{aligned}\text{Bias}\left\{\overline{I}_h(w;\mathbf{\Omega})\right\} &= \mathbf{E}\left\{\overline{I}_h(w;\mathbf{\Omega})\right\} - f(w) \\ &= \sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} \left(\lambda + \lambda^2\widetilde{f}(w') + e(w') - \lambda - \widetilde{f}(w)\right)|\mathbf{\Omega}|^{-1}W\left(\mathbf{\Omega}^{-1}\{w-w'\}\right) \\ &= \sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} \lambda^2\left\{\widetilde{f}(w') - \widetilde{f}(w) + e(w')\lambda^{-2}\right\}|\mathbf{\Omega}|^{-1}W\left(\mathbf{\Omega}^{-1}\{w-w'\}\right). \end{aligned} \tag{131}$$

To get an expression for the bias we implement a Taylor series, in the manner of e.g. Wand and Jones (1994), and so, assuming that at $w \neq 0$ that $\widetilde{f}(w)$ is twice differentiable then we have with $\widetilde{\mathbf{H}}_f(w)$ the Hessian of $\widetilde{f}(w)$, and $|w''-w'| < |w-w'|$

$$\widetilde{f}(w') = \widetilde{f}(w) + \nabla\widetilde{f}^T(w)(w'-w) + \frac{1}{2}(w'-w)^T\widetilde{\mathbf{H}}_f(w'')(w'-w).$$

Substituting this back into (131) we obtain the expression

$$\begin{aligned}\text{Bias}\left\{\overline{I}_h(w;\mathbf{\Omega})\right\} = &\sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} \lambda^2\left(\nabla\widetilde{f}^T(w)(w'-w) + \frac{1}{2}(w'-w)^T\widetilde{\mathbf{H}}_f(w'')(w'-w) + e(w')\lambda^{-2}\right) \\ &\cdot|\mathbf{\Omega}|^{-1}W\left(\mathbf{\Omega}^{-1}\{w-w'\}\right) \\ = &\frac{1}{2}\sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} \lambda^2\left((w'-w)^T\widetilde{\mathbf{H}}_f(k'')(w'-w) + e(w')\lambda^{-2}\right)|\mathbf{\Omega}|^{-1}W\left(\mathbf{\Omega}^{-1}\{w-w'\}\right).\end{aligned}$$

The linear term is eliminated because we either assume that W is isotropic or separable and even. We then see that only the diagonals matter (as the off-diagonals cancel due to the isotropy of the window). We define the ith diagonal term of the Hessian to be $\widetilde{f}_{ii}(w)$ and we write for the leading order term

$$|\text{Bias}_0\left\{\overline{I}_h(w;\mathbf{\Omega})\right\}| \leq \frac{\lambda^2}{2}\sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))}\sum_{i=1}^{d}\left\{\widetilde{f}_{ii}(w'')(k_i'-k_i)^2\right\}|\mathbf{H}|^{-1}W\left(\mathbf{H}^{-1}\{k-k'\}\right) + \max_{w\in\mathcal{K}(B_\square(\boldsymbol{l}))}\frac{1}{2}|e(w)|.$$

We can deduce directly by studying the error terms from Lemma 4.1 that $|e(w)| = \mathcal{O}\left(\frac{d}{\ell^{2-\beta}}\right) + \mathcal{O}\left(\frac{1}{\ell^{2\beta-1}}\right)$. Then we have

$$\begin{aligned}\left|\text{Bias}_0\left\{\overline{I}_h(w;\mathbf{\Omega})\right\}\right| \leq &\frac{\lambda^2}{2}\left|\sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))}\sum_{i=1}^{d}\left\{\widetilde{f}_{ii}(w'')(w_i'-w_i)^2\right\}|\mathbf{H}|^{-1}W\left(\mathbf{H}^{-1}\{w-w'\}\right)\right| + \mathcal{O}\left(\frac{d}{\ell^{2-\beta}}\right) \\ &+ \mathcal{O}\left(\frac{1}{\ell^{2\beta-1}}\right).\end{aligned}$$

Thus

$$\begin{aligned}\left|\text{Bias}_0\left\{\overline{I}_h(w;\mathbf{\Omega})\right\}\right| \leq &\frac{\lambda^2}{2}\max_i\max_{w''}|\widetilde{f}_{ii}(w'')|\sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))}\sum_{i=1}^{d}\left\{(w_i'-w_i)^2\right\}|\mathbf{H}|^{-1}W\left(\mathbf{H}^{-1}\{w-w'\}\right) \\ &+ \mathcal{O}\left(\frac{d}{\ell^{2-\beta}}\right) + \mathcal{O}\left(\frac{1}{\ell^{2\beta-1}}\right).\end{aligned} \tag{132}$$

We determine that the dominant latter term takes the form of

$$\sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))}\sum_{i=1}^{d}\left\{(w_i'-w_i)^2\right\}|\mathbf{H}|^{-1}W\left(\mathbf{H}^{-1}\{w-w'\}\right) \equiv \tau_h^2,$$

this in turn defining τ_h^2. Thus it follows:

$$\left|\text{Bias}_0\left\{\overline{I}_h(w;\mathbf{\Omega})\right\}\right| \leq \frac{\lambda^2}{2}\max_i|\widetilde{f}_{ii}(k')|\tau_h^2. \tag{133}$$

We shall now relate τ_h^2 to the parameters of the window. We start by assuming the wave number window is separable or

$$W(w) = \prod_{i=1}^{d} w(w_i). \tag{134}$$

We recall the constraints of producing an unbiased estimator, namely referring back to (52)

$$\sum_w W(w) = 1. \tag{135}$$

Once we renormalize then

$$\sum_{w\in\mathcal{K}} |\mathbf{H}|^{-1} W\left(\mathbf{H}^{-1}\{w - w'\}\right) \asymp \sum_w W(w) = 1. \tag{136}$$

For example if $w(\cdot)$ is the box window of width σ then with $k = p/\ell$:

$$w_i(w) = \left\{ \begin{array}{ll} \ell/\sigma & \text{if} \quad |w_i\ell/\sigma| < 1/2 \\ 0 & \text{o/w} \end{array} \right. . \tag{137}$$

We then see that of $\omega_{ii} = \omega$

$$\begin{aligned} \tau_\omega^2 &= \sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} (w')\cdot w' |\mathbf{\Omega}|^{-1} W\left(\mathbf{\Omega}^{-1} w'\right) \\ &= \sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} (w')\cdot w' \prod_i \omega_{ii}^{-1} \prod_{i=1}^d w\left(\frac{w_i'}{\omega_{ii}}\right). \end{aligned}$$

Thus we can compute

$$\begin{aligned} \tau_\omega^2 &= \sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} (w')\cdot w' \prod_i \ell(\sigma\omega_{ii})^{-1} \prod_{i=1}^d \mathbf{1}(|2w_i'\ell/(\sigma\omega_{ii})| < 1) \\ &= \sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} (w')\cdot w' \prod_j \ell(\sigma\omega_{jj})^{-1} \prod_{i=1}^d \mathbf{1}(|2w_i'\ell/(\sigma\omega_{ii})| < 1). \end{aligned}$$

Then it follows we can derive τ_ω^2 using a Riemann approximation to the sum. We then find:

$$\begin{aligned} \tau_\omega^2 &= \prod_{i=1}^d \int \sum_j (\ell^2\xi_j^2) \prod_i \left\{(\ell\sigma h_{ii})^{-1} \mathbf{1}(|2\xi_i/(\ell\sigma h_{ii})| < 1)\right\} \ell d\xi_i \\ &= \prod_{i=1}^d \int \cdots \int \sum_j (\ell^2\sigma^2 h_{jj}^2 \chi_j^2) \prod_i \{\mathbf{1}(|2\chi_i| < 1)\} \cdot d\chi_i \\ &= \sum_j (\ell^2\sigma^2 h_{jj}^2) \int_{-1/2}^{1/2} \cdots \int_{-1/2}^{1/2} \chi_j^2 d\chi_i \\ &= \sum_j (\ell^2\sigma^2 h_{jj}^2) \frac{1}{12} = d(\ell^2\sigma^2 h^2) \frac{1}{12}. \end{aligned}$$

O Proof of Lemma 6.2

For the periodogram we approximate the second order properties by

$$\begin{aligned} \mathbf{Var}\{I_0(w)\} &\sim f(w) - \lambda + \lambda^2 + O\left(|B|^{-1}\right) \\ \mathbf{Cov}\{I_0(w), I_0(w')\} &\sim O\left(|B|^{-1}\right). \end{aligned} \tag{138}$$

We derive these results from Theorem 6.2 combined with Proposition 6.2.

Starting from these results, we shall now calculate the variance of

$$\begin{aligned} \mathbf{Var}\left\{\widetilde{I}_h(w)\right\} &= \mathbf{Var}\left\{ \sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} I_0(w') |\mathbf{H}|^{-1} W\left(\mathbf{H}^{-1}(w' - w)\right) \right\} \\ &= \sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} \sum_{w''\in\mathcal{K}(B_\square(\boldsymbol{l}))} |\mathbf{H}|^{-1} W\left(\mathbf{H}^{-1}(w' - w)\right) |\mathbf{H}|^{-1} W\left(\mathbf{H}^{-1}(w'' - w)\right) \mathbf{Cov}\left\{I_0(w'), I_0(w'')\right\} \\ &= \sum_{k'\in\mathcal{K}(B_\square(\boldsymbol{l}))} |\mathbf{H}|^{-2} W^2\left(\mathbf{H}^{-1}(w' - w)\right) \left\{f(w') - \lambda + \lambda^2 + O\left(|B|^{-1}\right)\right\} + O\left(|B|^{-1}\right). \end{aligned} \tag{139}$$

Again using a separable and square window we have

$$\sum_{w'\in\mathcal{K}(B_\square(\boldsymbol{l}))} |\mathbf{H}|^{-2} W^2\left\{\mathbf{H}^{-1}(w' - w)\right\} = (\ell h\sigma)^{-1}. \tag{140}$$

P Proof of Theorem 6.1

We can now determine the mean square error. We have from combining Lemma 6.1 and Lemma 6.2 that (replacing $\ell h \sigma$ by $\tilde{\sigma}$)

$$\mathrm{MSE}\left\{\widetilde{I}_0(w_n;\mathbf{\Omega})\right\} \leq \frac{\lambda^4}{4}\max_i |\widetilde{f}_{ii}(w')|^2 \left\{d(\tilde{\sigma}^2)\frac{1}{12}\right\}^2 + \frac{f(w_n) - \lambda + \lambda^2 + O(|B|^{-1})}{\tilde{\sigma}} + O(|B|^{-1}).$$

Differentiating we then get

$$\begin{aligned}
\frac{\partial}{\partial\tilde{\sigma}}\mathrm{MSE}\left\{\widetilde{I}(w_n)\right\} &= \lambda^4 \max_i |\widetilde{f}_{ii}(w')|^2 \frac{d^2}{12^2}\tilde{\sigma}^3 - \frac{\lambda^2}{\tilde{\sigma}^2}\left\{1 + o(1)\right\} \\
\tilde{\sigma}^5 &= 12^2 \frac{\lambda^2}{d^2\lambda^4 \max_i |\widetilde{f}_{ii}(w')|^2} \\
\tilde{\sigma} &= \left(12^2 \frac{\lambda^2}{d^2\lambda^4 \max_i |\widetilde{f}_{ii}(w')|^2}\right)^{1/5} \\
\frac{\partial^2}{\partial\tilde{\sigma}^2}\mathrm{MSE}\left\{\widetilde{I}(w_n)\right\} &= \frac{\lambda^4 d^2}{4\times 12}\max_i |\widetilde{f}_{ii}(w')|^2\tilde{\sigma}^2 + 2\frac{\lambda^2}{\tilde{\sigma}^3} > 0.
\end{aligned}$$

We have therefore checked this is a minimum. We now need to estimate these parameters in order to get a suggested bandwidth. Some of the parameters are known, e.g. d. The curvature is a little bit more complex. We shall motivate our curvature estimate by a single length scale in the spectrum.

Q Proof of Proposition 7.1

Note that by direct computation

$$\begin{aligned}
\mathbf{E}\left\{I_D(w)\right\} &= \lambda + |B|^{-1}\int_{B^2} \rho^{(2)}(\|x-y\|)G(w\|x-y\|)\,dx\,dy \\
&= \lambda + |B|^{-1}\int |B\cap B_x|\,\rho^{(2)}(\|x\|)G(w\|x\|)\,dx \\
&= \lambda + |B|^{-1}\int_0^\infty \rho^{(2)}(r)r^{d-1}G(wr)\int_{S^{d-1}} |B\cap B_{ru}|\,du\,dr \\
&= \lambda + d\cdot|B|^{-1}\int_0^\infty \rho^{(2)}(r)r^{d-1}G(wr)\overline{\nu}_B(r)\,dr \\
&= \lambda + d\,|B|^{-1}\int_0^\infty \gamma(r)r^{d-1}G(wr)\overline{\nu}_B(r)\,dr + \lambda^2 d\cdot b_d\,|B|^{-1}\,H(w).
\end{aligned}$$

R Proof of Proposition 7.2

An interesting question is naturally; what if we use Diggle's estimator, even if the point process is not isotropic. We recall that the estimator takes the form

$$I_D(|w|) = \hat{\lambda} + |B|^{-1}\sum_{x\neq y} J_0\left(2\pi|w|\cdot\|x-y\|\right). \tag{141}$$

We can still compute the estimator for any observed point-process X, even if X was not generated by an isotropic process. The estimator $I_D(|w|)$ has expectation

$$\begin{aligned}
\mathbf{E}I_D(|w|) &= \lambda + |B|^{-1}\sum_{x\neq y}\mathbf{E}J_0\left(2\pi|w|\cdot\|x-y\|\right) &(142)\\
&= \lambda + |B|^{-1}\int_B\int_{B_{-x}} J_0\left(2\pi|w|\cdot|z|\right)\rho^{(2)}(z)\,dz\,dx \\
&= \lambda + |B|^{-1}\int_B\int_{B_{-x}} J_0\left(2\pi|w|\cdot|z|\right)\left\{\rho^{(2)}(z)-\lambda^2\right\}\,dz\,dx + \lambda^2|B|^{-1}\int_B\int_{B_{-x}} J_0\left(2\pi|w|\cdot|z|\right)\,dz\,dx \\
&= \lambda + |B|^{-1}\int_{\mathbb{R}^d} |B\times B_{-z}|\cdot J_0\left(2\pi|w|\cdot|z|\right)\left\{\rho^{(2)}(z)-\lambda^2\right\}\,dz + \lambda^2|B|^{-1}\int_B |B\times B_{-z}|\cdot J_0\left(2\pi|w|\cdot|z|\right)\,dz \\
&= \mathbf{E}I_D^{(1)}(|w|) + \mathbf{E}I_D^{(2)}(|w|) + \mathbf{E}I_D^{(3)}(|w|), &(143)
\end{aligned}$$

the latter defining the form of these three contributions.

Just like before we shall explicitly demonstrate the effects of this convolution. We have that the Fourier transform if the Bessel function is

$$\begin{aligned}\mathcal{F}\left\{J_0\left(2\pi|w|\cdot|z|\right)\right\}(u) &= \iint J_0\left(2\pi|w|\cdot|z|\right)e^{-2\pi i z^T u}dz \\ &= \int_0^\infty \int_0^{2\pi} J_0\left(2\pi|w|r\right)e^{-2\pi i r|u|\cos(\phi-\phi_u)}2\pi|w|d|w|d\phi \\ &= \int_0^\infty J_0\left(2\pi|w|r\right)2\pi|w|d|w|\int_0^{2\pi} e^{-2\pi i|w||u|\cos(\phi-\phi_u)}d\phi. \end{aligned} \tag{144}$$

We now note that

$$J_0(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ix\sin t}dt.$$

Thus

$$\mathcal{F}\left\{J_0\left(2\pi|w|\cdot|z|\right)\right\}(u) = \int_0^\infty J_0\left(2\pi|w|r\right)(2\pi)^2 J_0\left(2\pi r|u|\right)\ rdr = (2\pi)^2\frac{\delta(|w|-|u|)}{|w|}.$$

We now use the convolution theorem to deduce that:

$$\begin{aligned}\mathbf{E}I_D^{(3)}(|w|) &= |B|^{-1}\int_B |B\times B_{-z}|\cdot J_0\left(2\pi|w||z|\right)\ dz \\ &= \int_{\mathbb{R}^d}\frac{\delta(|w|-|u|)}{k}|B|^{-1}T(B,u)\ du. \end{aligned} \tag{145}$$

Thus the low frequency bias is determined by this term. Thus the bias is worse for this version of the wave number domain representation. We then see how this "blurs out" the sinc kernel. This implies that the second part of the integral becomes

$$\mathbf{E}I_D^{(2)}(|w|) = |B|^{-1}\int_{\mathbb{R}^d}|B\times B_{-z}|\cdot J_0\left(2\pi|w||z|\right)\left\{\rho^{(2)}(z)-\lambda^2\right\}\ dz.$$

We therefore need to compute

$$\begin{aligned}\mathcal{F}\left\{|B|^{-1}\cdot|B\times B_{-z}|\cdot J_0\left(2\pi|w||z|\right)\right\}(u) &= \int_{\mathbb{R}^d}|B|^{-1}\cdot|B\times B_{-z}|\cdot J_0\left(2\pi|w||z|\right)e^{-2\pi i u^T z}\ dz \\ &= \int_{\mathbb{R}^d}\frac{\delta(|w|-|u'|)}{|w|}|B|^{-1}T(B,u'-u)\ du' \\ \mathcal{F}\left\{\rho^{(2)}(z)-\lambda^2\right\}(u) &= \lambda^2\widetilde{f}(u). \end{aligned} \tag{146}$$

With these pieces we have that

$$\mathbf{E}I_D^{(2)}(|w|) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\lambda^2\widetilde{f}(u')\frac{\delta(|w|-|u''|)}{|k|}|B|^{-1}T(B,u''-u')\ du''du'. \tag{147}$$

We now see that further blurring is present in (147). In the standard non-isotropic case this was just a convolution of $\widetilde{f}(w)$ with $|B|^{-1}T(B,w)$. To get a feeling for its behaviour we note that as $|B|^{-1}T(B,w)\to\delta(w)/(2\pi)$. In this case we get for $d=2$

$$\begin{aligned}\mathbf{E}I_D^{(2)}(|w|) &\to \int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\lambda^2\widetilde{f}(u')\frac{\delta(|w|-|u''|)}{2\pi|w|}\delta(u''-u')\ du''du' \\ &= \int_{\mathbb{R}^2}\lambda^2\widetilde{f}(u'')\frac{\delta(|w|-|u''|)}{2\pi|w|}\ du'' \\ &= \lambda^2\int_{\mathbb{R}^+}\widetilde{f}(u'')\frac{\delta(|w|-|u''|)}{|w|}|u''|\ d|u''| \\ &= \widetilde{f}(|w|). \end{aligned} \tag{148}$$

This shows that asymptotically we would recover the isotropic spectral density from this component. Note that the first term, or $\mathbf{E}I_D^{(2)}(|w|)$ is just λ and already determined. By combining the three terms, we are left with the expectation of the Diggle isotropic estimator. By combining the terms with arrive at the result stated in the proposition.