Geophys. J. Int. (2026) , http://dx.doi.org/10.1093/gji/

Maximum-likelihood estimation of the Matérn covariance structure of isotropic spatial random fields on finite, sampled grids

Frederik J. Simons[1,2], Olivia L. Walbert[1], Arthur P. Guillaumin[3], Gabriel L. Eggers[4], Kevin W. Lewis[5], and Sofia C. Olhede[6]

[1] *Department of Geosciences, Princeton University, Princeton, NJ 08544, USA. E-mail: fjsimons@alum.mit.edu*
[2] *Program in Applied & Computational Mathematics, Princeton University, Princeton, NJ 08544, USA*
[3] *Queen Mary University of London, E1 4NS, London, UK*
[4] *Wesleyan University, Middletown, CT 06459, USA*
[5] *Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, USA*
[6] *Ecole Polytechnique Fédérale de Lausanne, Switzerland*

26 January 2026

SUMMARY

We present a statistically and computationally efficient spectral-domain maximum-likelihood procedure to solve for the structure of Gaussian spatial random fields within the Matérn covariance hyperclass. For univariate, stationary, and isotropic fields, the three controlling parameters are the process variance, smoothness, and range. The debiased Whittle likelihood maximization explicitly treats discretization and edge effects for finite sampled regions in parameter estimation and uncertainty quantification. As even the 'best' parameter estimate may not be 'good enough', we provide a test for whether the model specification itself warrants rejection. Our results are practical and relevant for the study of a variety of geophysical fields, and for spatial interpolation, out-of-sample extension, kriging, machine learning, and feature detection of geological data. We present procedural details and high-level results on real-world examples.

Key words: Fourier transforms. Geostatistics. Spectral analysis. Statistical methods.

1 INTRODUCTION

What numbers, which statistical notions capture the "essence" of a spatial patch of geophysical data? If it were a stationary, white, Gaussian process, simply, its population mean and variance would be sufficient. However, computed over differing window sizes, or at varying resolution, sample means and variances of geophysical data sets fluctuate non-erratically, hence "whiteness" immediately proves to be an untenable assumption. Reporting summary statistics over changing baselines (e.g., Sharpton & Head 1985; Aharonson et al. 1998; Grohmann et al. 2011; Rosenburg et al. 2011, for Earth and planetary topography) effectively subscribes to the data as realizations from a spatially *correlated* (locally stationary, Gaussian) process. Estimating even just the variance of said field (that is, at a point), with little bias (that is, accurately) and with a reasonable estimation variance (that is, precisely), requires knowledge of the co-variance (between pairs of points), and an acknowledgment of the inherently finite nature of the available spatial patch over which a single (geological) process can be thought of as representative. We are led to the estimation of (the parameters of) a spatial covariance function, or alternatively and equivalently, of a spectral density function, from (hopefully locally) stationary sampled data that are (irregularly) bounded (see, e.g., Valentine & Davies 2020). The first (spatial estimation) is notoriously noisy since it requires finding data pairs at increasing offsets, and computationally time-consuming when it entails inverting a covariance matrix (see, e.g., Mardia & Marshall 1984; Kitanidis & Lane 1985; Vecchia 1988; Grainger et al. 2021). The second (spectral estimation), while typically faster and less noisy, is famously affected by aliasing, finite-field, and edge effects (e.g., Stein 1995; Appourchaux et al. 1998a,b; Hamilton 2009a,b). Fourier-domain artifacts lead to estimation bias especially for multidimensional data sets, where the nefarious influence of boundary terms dominates the mean-squared error (Dahlhaus & Künsch 1987).

Estimating model parameters of random fields typically relies on spatial (Kent & Mardia 1996) or spectral (Whittle 1953) likelihood maximization. If the former is often slow to compute, and the latter suffers from potential boundary effects, there are notable modifications to both. For the spatial methods, Vecchia approximations to Gaussian log-likelihoods are fast and accurate (see, e.g., Katzfuss & Guinness 2021; Porcu et al. 2024). Circulant-embedding based methodologies (Guinness & Fuentes 2017; Guinness 2019) effectively recompute the Fourier basis rather than assuming it known from homogeneity in space. Vecchia-type approaches do require a local neighborhood for prediction, chosen appropriately. Some of these methods are discussed relative to spectral approaches both in theory and via simulations by Guillaumin

et al. (2022), who furthermore introduce adjustments to correct for partial-grid and boundary effects with the spectral methods, developing a large-sample theory for the appropriate asymptotic framework, upon which we build in this paper.

Common choices for parameterized covariance functions of Gaussian random fields are exponential or squared-exponential forms (e.g., Tarantola & Nercessian 1984; Montagner 1986; Guðmundsson et al. 1990; Baig et al. 2003; Baig & Dahlen 2004), defined solely by a variance and a correlation range. Both are special cases of the Matérn (1960) class of covariance functions (Guttorp & Gneiting 2006), with a specific smoothness or mean-squared differentiability (Adler 1981; Christakos 1992), itself an inversion parameter of interest in a multitude of geophysical application domains such as seismology (e.g., Wu & Aki 1985; Wu & Flatté 1990; Becker et al. 2007; Carpentier & Roy-Chowdhury 2007), seafloor bathymetry and oceanography (Goff & Jordan 1989a,b; Goff & Arbic 2010; Sandwell et al. 2022; Simon et al. 2026), helioseismology (e.g., Gizon & Birch 2004), hydrology (e.g., Rodríguez-Iturbe & Mejía 1974; Kitanidis & Lane 1985), meteorology, and climate science (e.g., Handcock & Wallis 1994; Paciorek & Schervish 2006; Lindgren et al. 2011; North et al. 2011; Sun et al. 2015).

In this paper we present a univariate spectral-domain debiased "Whittle" maximum-likelihood procedure (Simons & Olhede 2013; Guillaumin et al. 2017; Sykulski et al. 2019) that estimates the variance, smoothness, and range of an isotropic Matérn Gaussian process, from sampled spatial data. We show how to obtain unbiased estimates for these size, scale, and shape parameters when the region studied is neither rectangular nor circular, nor completely sampled. We exactly calculate their estimation covariance, correctly blurred for finite-field effects and without neglecting commonly omitted wave vector correlation effects, such that the results from differently sized and sampled patches can be compared robustly. We develop and comprehensively illustrate our algorithms for completely sampled rectangular data sets.

While process variance estimation necessitates recognizing the effects of spatial covariance, which is aided by the ability to extend the size of the observation domain, intuitively "beyond the reach of the range", conversely, to characterize the smoothness of a process requires neighboring observations that are highly correlated. In other words, both increasing (growing) domain and fixed domain (infill) asymptotics (and mixed-domain, see Chu 2023) are relevant considerations (see, e.g., Zhang 2004; Zhang & Zimmerman 2005) that we are, however, unable to fully address in this paper, although the computer code that we freely distribute (see Sec. 7) is designed to give the reader all the tools in hand to be able to do so. Neither form of asymptotics is a panacea, and, furthermore, an unfortunate aspect of the Matérn class is that in two dimensions, neither the variance nor the range are consistently estimable under fixed-domain asymptotics, although there is a non-linear function of them and the smoothness that is (see Stein 1995). Guillaumin et al. (2022), who provided the theoretical basis for our work, introduced the notion of *significant correlation contribution* as a means of capturing whether a covariance function under a given spatial sampling mechanism allows us to distinguish parameter values. It cannot eliminate intrinsic problems with the flexibility of the Matérn family, and in some scenarios it may indeed be beneficial to *fix* the smoothness. While the full theoretical framework allows us to determine when we are able to estimate a specific Matérn family with a given sampling mechanism, it is our goal, with this paper, to provide sufficient, but not exhaustive, guidance, accompanied by computer code, to make Matérn covariance estimation practical and sensible for geoscientists.

Our results have widespread implications for the study of geophysical fields, and should be interpreted in the light of our trying to derive "process" from "parameters", the end goal being to be able to assign likely formation mechanisms and histories for the patches under consideration. Our results should also be relevant for whomever needs to perform spatial interpolation or out-of-sample extension (e.g., via kriging) on geological data (Journel & Huijbregts 1978; Christakos 1992; Cressie 1993; Stein 1999). They also carry consequences for machine learning and feature detection (Rasmussen & Williams 2006; Porcu et al. 2024). We present procedural details but also focus on high-level results that have real-world applications. We illustrate our methodology on four geologically and geophysically relevant data sets, assuming stationarity within patches that have been selected via user interpretation.

2 PRELIMINARIES

Readers wishing to come to terms with the geological, geophysical, and geodetic definitions of 'relief', 'topography', or 'elevation' are directed to Lambeck (1988), Hofmann-Wellenhof & Moritz (2006), and Wieczorek (2015). To make the jump from geology and geophysics to statistics, particularly in this context, we first and foremost recommend (re)reading Goff & Jordan (1988, 1989a), who also discuss anisotropic processes. The material in this section is both an extension and a specialization of the multivariate results of Simons & Olhede (2013), which is to be consulted for further details. Here, we use a more explicit notation, adapt some of the normalizations, and make a number of modifications—but most importantly, we restrict our analysis to univariate two-dimensional Cartesian isotropic Gaussian fields.

2.1 Continuous framework

Here we draw most heavily on Sections 2.1 and 4.1 of Simons & Olhede (2013). Referring furthermore to Percival & Walden (1993), Stein (1999), and Vanmarcke (2010) for additional considerations and terminology, to Abramowitz & Stegun (1965) and Gradshteyn & Ryzhik (2000) for properties of special functions, we begin by defining the particular quantities of interest in the *spatial* and the *spectral* domains.

2.1.1 Stationarity

A geophysical field $\mathcal{H}(\mathbf{x})$ is considered to be a zero-mean, finite-variance, stationary, two-dimensional random field with finite second moments. Under what is known as the Cramér (1942) representation, there exists a spectral orthogonal-increment process, $d\mathcal{H}(\mathbf{k})$, according

to which the spatial field

$$\mathcal{H}(\mathbf{x}) = \iint e^{i\mathbf{k}\cdot\mathbf{x}}\, d\mathcal{H}(\mathbf{k}), \qquad \mathbf{x} \in \mathbb{R}^2. \tag{1}$$

The integration is over the space containing all wave vectors $\mathbf{k}$. In the case of strict bandlimitation or very fast spectral decay we may restrict the computations to the Nyquist plane $[-\pi, \pi] \times [-\pi, \pi]$. The expectation of $d\mathcal{H}(\mathbf{k})$ over many realized fields,

$$\langle d\mathcal{H}(\mathbf{k}) \rangle = 0, \tag{2}$$

and its variance, in the absence of covariance between wave vectors, defines a power-spectral density, $\mathcal{S}(\mathbf{k})$, in the form of the expectation

$$\langle d\mathcal{H}(\mathbf{k})\, d\mathcal{H}^*(\mathbf{k}') \rangle = \mathcal{S}(\mathbf{k})\, d\mathbf{k}\, d\mathbf{k}'\, \delta(\mathbf{k}, \mathbf{k}'), \tag{3}$$

where $\delta(\mathbf{k}, \mathbf{k}')$ is the Dirac delta function. When eqs (1)–(3) hold, the spatial auto-covariance, $\mathcal{C}(\mathbf{x}, \mathbf{x}')$, displays stationarity by being dependent on *separation*, $\mathbf{x} - \mathbf{x}'$, only, since in that case we can write for the expectation of the two-point spatial-domain products

$$\langle \mathcal{H}(\mathbf{x}) \mathcal{H}^*(\mathbf{x}') \rangle = \iint e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} \mathcal{S}(\mathbf{k})\, d\mathbf{k} = \mathcal{C}(\mathbf{x} - \mathbf{x}'). \tag{4}$$

The spectral variance (at the wave vectors $\mathbf{k}$) and the spatial covariance (in the *lag* variables $\mathbf{x}$) form a Wiener-Khintchine Fourier pair,

$$\mathcal{C}(\mathbf{x}) = \iint e^{i\mathbf{k}\cdot\mathbf{x}} S(\mathbf{k})\, d\mathbf{k}, \tag{5}$$

$$\mathcal{S}(\mathbf{k}) = \frac{1}{(2\pi)^2} \iint \mathcal{C}(\mathbf{x})\, e^{-i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{x}. \tag{6}$$

The zero-wavenumber intercept of the spectral density is the zeroth moment of the spatial covariance:

$$\mathcal{S}(0) = \frac{1}{(2\pi)^2} \iint \mathcal{C}(\mathbf{x})\, d\mathbf{x}. \tag{7}$$

2.1.2 Isotropy

Under isotropy, abusing notation, $\mathcal{S}(\mathbf{k}) = \mathcal{S}(k)$, depending only on the scalar wavenumber $k = \|\mathbf{k}\|$. Integrating over the polar angles to bring out J_0, the Bessel function of the first kind and of order zero, the spatial covariance,

$$\langle \mathcal{H}(\mathbf{x}) \mathcal{H}^*(\mathbf{x}') \rangle = 2\pi \int J_0(k\|\mathbf{x} - \mathbf{x}'\|)\, \mathcal{S}(k)\, k\, dk = \mathcal{C}(\|\mathbf{x} - \mathbf{x}'\|), \tag{8}$$

is dependent only on *distance*, $\|\mathbf{x} - \mathbf{x}'\|$, not direction. Since $J_0(0) = 1$, the isotropic spatial variance is then given by

$$\langle \mathcal{H}(\mathbf{x}) \mathcal{H}^*(\mathbf{x}) \rangle = 2\pi \int \mathcal{S}(k)\, k\, dk = \mathcal{C}(0) = \sigma^2. \tag{9}$$

Introducing the distance variable r and integrating over the angular polar coordinate, we rewrite eq. (7) as

$$\mathcal{S}(0) = \frac{1}{2\pi} \int \mathcal{C}(r)\, r\, dr. \tag{10}$$

We follow Vanmarcke (2010) in adopting the term 'fluctuation scale' for $\mathcal{S}(0)/\sigma^2$, the mass of the spatial *correlation* function $\mathcal{C}(r)/\sigma^2$.

Isotropy remains a restrictive—but testable—assumption, which we will be relaxing in forthcoming work. Any future discussion of anisotropy will entail evaluating it against the null-hypothesis of isotropic behavior, in the possible presence of anisotropic sampling patterns, for which the present paper provides all the necessary statistical machinery.

2.1.3 Matérnity

We further specify the field as a member of the Matérn class (Stein 1999), which is very general and widely applicable to processes with monotonically decreasing autocovariance (Guttorp & Gneiting 2006). The isotropic d-dimensional Matérn spectral density $\mathcal{S}^d_{\boldsymbol{\theta}}(k)$ assumes the parameterized form (Handcock & Wallis 1994; Wang et al. 2023)

$$\mathcal{S}^d_{\boldsymbol{\theta}}(k) = \frac{\Gamma(\nu + d/2)}{\Gamma(\nu)} \frac{\sigma^2}{\pi^{d/2}} \left(\frac{4\nu}{\pi^2\rho^2}\right)^{\nu} \left(\frac{4\nu}{\pi^2\rho^2} + k^2\right)^{-\nu-d/2}, \tag{11}$$

where Γ is the gamma function, and which, in two dimensions, $d = 2$, as we subsumed earlier and maintain from now on, specifies to

$$\mathcal{S}_{\boldsymbol{\theta}}(k) = \sigma^2 \frac{\pi\rho^2}{4} \left(\frac{4\nu}{\pi^2\rho^2}\right)^{\nu+1} \left(\frac{4\nu}{\pi^2\rho^2} + k^2\right)^{-\nu-1}. \tag{12}$$

With this model, our principal unknowns are its three strictly positive parameters, denoted generically as $\theta > 0$, which we collect in the set

$$\boldsymbol{\theta} = [\sigma^2 \;\; \nu \;\; \rho\,]^T. \tag{13}$$

The 'variance', σ^2, indeed satisfies eq. (9) upon substitution with eq. (12). At short wavelengths, when k is large, the spectrum $\mathcal{S}_{\boldsymbol{\theta}}(k)$ decays at a rate that depends on the 'smoothness', ν, which expresses the $\lceil \nu - 1 \rceil$ times (mean-squared) 'differentiability' of the process (Handcock & Stein 1993). The behavior at the longest wavelengths, for small k, is controlled by the combined effect of σ^2 and ρ. The fluctuation scale

$$\frac{\mathcal{S}_{\boldsymbol{\theta}}(0)}{\sigma^2} = \frac{\pi\rho^2}{4}. \tag{14}$$

The isotropic Matérn spatial covariance $\mathcal{C}_{\boldsymbol{\theta}}(r)$, in terms of the lag distance r, is unlike its spectral counterpart (11) in requiring no dimensional specification, that is, independently of d,

$$\mathcal{C}_{\boldsymbol{\theta}}(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{2\nu^{\frac{1}{2}}}{\pi\rho} r \right)^{\nu} K_\nu \left(\frac{2\nu^{\frac{1}{2}}}{\pi\rho} r \right), \tag{15}$$

with K_ν the modified Bessel function of the second kind. The asymptotic behavior $K_\nu(z) \to \Gamma(\nu)\,(z/2)^{-\nu}/2$ for small z, verifies that $\mathcal{C}_{\boldsymbol{\theta}}(0) = \sigma^2$ as in eq. (9). For low values of ν, furthermore, $\mathcal{C}_{\boldsymbol{\theta}}(\pi\rho) \approx \sigma^2/3$. In other words, spatial correlations generally die down by a factor of about two-thirds at distances beyond $r \approx \pi\rho$, hence the name for the third parameter, the 'correlation length' or 'range', ρ.

The power accumulated over a certain wavenumber interval, counting from the origin, is given by the distribution function

$$\mathcal{P}_{\boldsymbol{\theta}}(k) = 2\pi \int_0^k \mathcal{S}_{\boldsymbol{\theta}}(k')\, k'\, dk' = \sigma^2 \left[1 - \left(\frac{4\nu}{\pi^2\rho^2} \right)^{\nu} \left(\frac{4\nu}{\pi^2\rho^2} + k^2 \right)^{-\nu} \right]. \tag{16}$$

As expected $\mathcal{P}_{\boldsymbol{\theta}}(0) = 0$ and $\mathcal{P}_{\boldsymbol{\theta}}(\infty) = \sigma^2$. We define the wavenumbers k_α at which the power reaches $100 \times \alpha$ per cent of the total,

$$\mathcal{P}_{\boldsymbol{\theta}}(k_\alpha) = \alpha\sigma^2, \tag{17}$$

which, from eq. (16), is solved analytically by

$$k_\alpha = \frac{2\nu^{\frac{1}{2}}}{\pi\rho} \left[(1-\alpha)^{-1/\nu} - 1 \right]^{\frac{1}{2}}. \tag{18}$$

It can be readily verified that $k_0 = 0$ and $k_1 = \infty$. For convenience, we express the equivalent wavelengths in the notation $\lambda_{100\alpha} = 2\pi/k_\alpha$.

The flexible generality of the Matérn class is appreciated by evaluating the correlation functions for special values of ν (Guttorp & Gneiting 2006). Notably, when $\nu = 1/2$, the correlation function decays exponentially, and when $\nu \to \infty$, as a Gaussian—a squared exponential. Other examples include the Von Kármán ($\nu = 1/3$), Whittle ($\nu = 1$), and second-order ($\nu = 3/2$) and third-order ($\nu = 5/2$) autoregressive correlation models. Despite all of its generality and wide applicability, the exponentially decaying, isotropic, three-parameter Matérn model cannot capture all random fields under study. It is not identifiable with (multi-) fractal, scale-invariant, self-affine, or self-similar behavior (see, e.g., Mareschal 1989; Herzfeld et al. 1995; Gneiting et al. 2012; Landais et al. 2019), as it is self-affine only at high wavenumber. For temporal processes, Lilly et al. (2017) discuss when the Matérn process is more appropriate to use than fractional Brownian motion. Developing a general set of methods outside these constraints, e.g., when the spectrum is no longer bounded (Reed et al. 2002), remains outstanding as future work.

Fig. 1 provides intuitive insight into the role that the three parameters σ^2, ν, and ρ play in the spatial behavior of Matérn random fields, synthesized by the circulant embedding procedure outlined in the next section.

2.2 Lattice framework

Here we rely mostly on sections 2.1, 4.2, and A6 of Simons & Olhede (2013). The properties of the finite and sampled, i.e., *windowed* discrete processes, as will be experienced in computational data analysis, differ markedly from the behavior of the idealized, infinite, continuous models discussed in the previous section, and those two viewpoints need to be explicitly reconciled.

2.2.1 Discretization

For simplicity of notation, $\mathbf{x}$ now maps out a rectangular $K = M \times N$ grid of "pixels" with spatial extent Δx and Δy given by

$$\mathbf{x} = \big\{ (m\,\Delta x, n\,\Delta y) \big\}, \qquad \text{for} \qquad m = 0, \ldots, M-1 \qquad \text{and} \qquad n = 0, \ldots, N-1. \tag{19}$$

We define the discrete Fourier transform of the noiseless measurements of the spatial process $\mathcal{H}(\mathbf{x})$ obtained after sampling as

$$H(\mathbf{k}) \equiv \frac{1}{2\pi} \left(\frac{\Delta x \Delta y}{MN} \right)^{\frac{1}{2}} \sum_{\mathbf{x}} \mathcal{H}(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}. \tag{20}$$

Sampled in spectral space, the finite set of wave vectors is now, with m and n as in eq. (19),

$$\mathbf{k} = \big(k_x, k_y\big) = \left\{ \left(\frac{2\pi}{M\Delta x} \left[-\left\lfloor \frac{M}{2} \right\rfloor + m \right],\ \frac{2\pi}{N\Delta y} \left[-\left\lfloor \frac{N}{2} \right\rfloor + n \right] \right) \right\}, \tag{21}$$

and on this complete Nyquist grid we now identify eq. (1), consistently with eq. (20), with

$$\mathcal{H}(\mathbf{x}) \equiv \frac{2\pi}{(MN\Delta x \Delta y)^{\frac{1}{2}}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} H(\mathbf{k}). \tag{22}$$

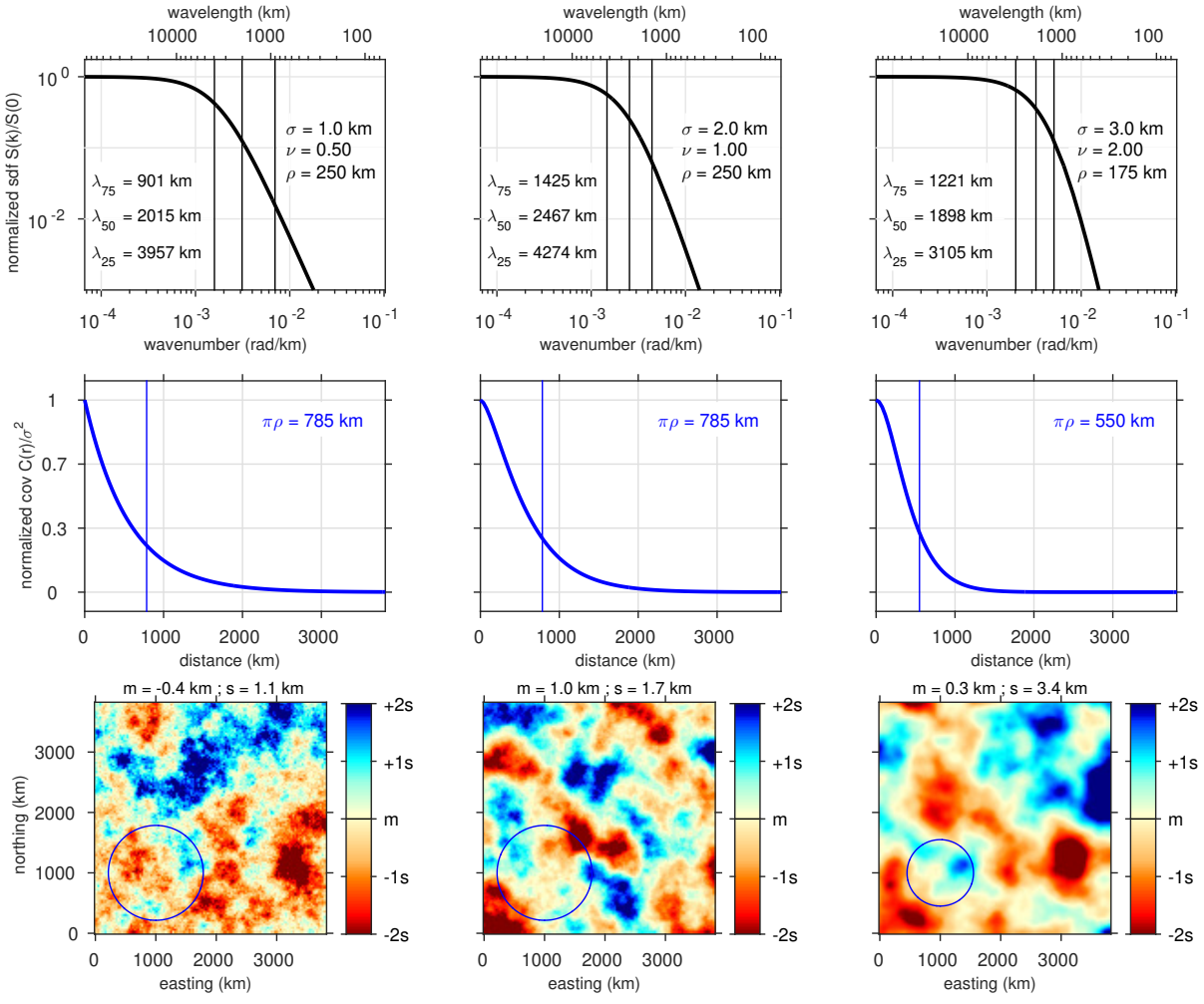

Figure 1. Random fields generated from stationary isotropic Matérn models with variances σ^2, differentiabilities ν, and correlation lengths ρ, as indicated. (*Top:*) Normalized spectral densities, $\mathcal{S}_{\boldsymbol{\theta}}(k)/\mathcal{S}_{\boldsymbol{\theta}}(0)$, from eq. (12). The vertical black lines identify the wavenumbers k_α at which the power reaches $100 \times \alpha$ per cent of the variance, from eq. (18), in wavelengths $\lambda_{100\alpha} = 2\pi/k_\alpha$, as labeled. (*Middle:*) Correlations, the normalized spatial covariances, $\mathcal{C}_{\boldsymbol{\theta}}(r)/\sigma^2$, from eq. (15). The vertical blue lines are drawn at the values $\pi\rho$, the distances at which the correlations die down to approximately one third of the variance. (*Bottom:*) Field realizations. The blue circles have radii $\pi\rho$, drawn for visual guidance. In the titles, m and s identify the sample means and standard deviations.

2.2.2 Blurring

Obtaining space-domain realizations from a population of random fields specified by a certain spectral density such as given by eq. (12) is possible by generating Fourier coefficients $H(\mathbf{k})$, as in eq. (20), directly on the spectral grid (21), and by inverse Fourier transformation, as in eq. (22), onto the spatial grid (19). These $H(\mathbf{k})$ should be drawn from a zero-mean complex proper Gaussian distribution, with expectation zero, $\langle H(\mathbf{k})\rangle = 0$, and with a covariance $\langle H(\mathbf{k})H^*(\mathbf{k})\rangle$ that will be influenced by the chosen size, shape and discretization of the region under consideration; i.e., it will be different from the theoretical quantity $\langle d\mathcal{H}(\mathbf{k})\, d\mathcal{H}^*(\mathbf{k}')\rangle$ of eq. (3), which involved the true density $\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k})$.

Simons & Olhede (2013) showed the covariance of a finite set of gridded Fourier coefficients can at best offer a blurred and correlated version of the true spectral variance (see also Fournier et al. 2014). To us, their eq. (9) reads, using $\mathcal{S}_{\boldsymbol{\theta}}(k) = \mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k})$ from our eq. (12),

$$\langle H(\mathbf{k})H^*(\mathbf{k}')\rangle = \iint_{\mathbf{k}} D_K(\mathbf{k}-\mathbf{k}'')D_K^*(\mathbf{k}'-\mathbf{k}'')\,\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k}'')\,d\mathbf{k}'' \quad \text{with} \quad D_K(\mathbf{k}) = \frac{1}{2\pi}\left(\frac{\Delta x\Delta y}{MN}\right)^{\frac{1}{2}} \sum_{\mathbf{x}} e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{23}$$

whereby $D_K(\mathbf{k})$ is the Dirichlet kernel, the suitably normalized discrete Fourier transform of the unit sampling operator. They simulated fields by incorporating the blurring but ignoring the correlation, that is, following their eqs (12) and (83), they approximated eq. (23) as

$$\langle H(\mathbf{k})H^*(\mathbf{k}')\rangle \approx \delta(\mathbf{k},\mathbf{k}') \iint_{\mathbf{k}} \left|D_K(\mathbf{k}-\mathbf{k}')\right|^2 \mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k}')\,d\mathbf{k}' = \bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})\delta(\mathbf{k},\mathbf{k}'). \tag{24}$$

The blurred spectral density $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$ is the convolution of the true spectral density $\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k})$ with the Fejér kernel $|D_K(\mathbf{k}-\mathbf{k}')|^2$, which embodies the effects both of sampling and finite domain size, and reduces to the Dirac delta for continuous processes on infinite domains. Simons & Olhede (2013) carried out the blurring *approximately*, via grid refinement, convolution, and subsampling.

Guillaumin et al. (2022), in contrast, in their Lemmata 1 and 2, showed how to *exactly* incorporate the spectral blurring effect of applying arbitrary data windows (see also Fuentes 2007), including irregular boundaries and incomplete sampling, at a much reduced computational cost. We rewrite the discrete Fourier transform in eq. (20), which silently selected samples on the grid $\mathbf{x}$ but did not modify or weight them, to incorporate a non-constant, unit-normalized, data window, $w(\mathbf{x})$, as follows:

$$H(\mathbf{k}) \equiv \frac{1}{2\pi}\left(\frac{\Delta x \Delta y}{MN}\right)^{\frac{1}{2}} \sum_{\mathbf{x}} w(\mathbf{x})\mathcal{H}(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}. \tag{25}$$

Using the definition in eq. (4), the sample variance of the windowed Fourier coefficients is

$$\text{var}\,\{H(\mathbf{k})\} = \frac{1}{(2\pi)^2}\left(\frac{\Delta x \Delta y}{MN}\right) \sum_{\mathbf{x}}\sum_{\mathbf{x}'} w(\mathbf{x})w(\mathbf{x}')\mathcal{C}_{\boldsymbol{\theta}}\left(\mathbf{x}-\mathbf{x}'\right) e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}, \tag{26}$$

$$= \frac{1}{(2\pi)^2}\left(\frac{\Delta x \Delta y}{MN}\right) \sum_{\mathbf{y}}\left(\sum_{\mathbf{x}-\mathbf{y}\,\cap\,\mathbf{x}} w(\mathbf{x})w(\mathbf{x}-\mathbf{y})\right) \mathcal{C}_{\boldsymbol{\theta}}\left(\mathbf{y}\right) e^{-i\mathbf{k}\cdot\mathbf{y}}, \tag{27}$$

following a change of variables and a change in the order of summation, noting that the first sum is over the separation grid

$$\mathbf{y} = \big\{(m'\Delta x, n'\Delta y)\big\}, \qquad \text{with the mirrored index sets} \qquad m' = -M+1,\ldots,M-1 \quad \text{and} \quad n' = -N+1,\ldots,N-1, \tag{28}$$

and the second sum, for each element of $\mathbf{y}$, over the subset of $\mathbf{x}-\mathbf{y}$ that belongs to the original grid $\mathbf{x}$, so as to stay within the original integration domain. Our manipulations allow us to isolate and sum out the interior term, which we rewrite more explicitly as

$$W(\mathbf{y}) = \sum_{\max(\mathbf{x}-\mathbf{y},|\mathbf{y}|)} w(\mathbf{x})w(\mathbf{x}-|\mathbf{y}|) = \sum_{\min(\mathbf{x}-\mathbf{y},|\mathbf{y}|)} w(\mathbf{x})w(\mathbf{x}+|\mathbf{y}|). \tag{29}$$

Reviewing what this implies for the variance of the Fourier coefficients of a windowed and sampled field $\mathcal{H}(\mathbf{x})$, we rewrite eqs (26)–(27) as

$$\text{var}\,\{H(\mathbf{k})\} = \frac{1}{(2\pi)^2}\left(\frac{\Delta x \Delta y}{MN}\right) \sum_{\mathbf{y}} W(\mathbf{y})\mathcal{C}_{\boldsymbol{\theta}}\left(\mathbf{y}\right) e^{-i\mathbf{k}\cdot\mathbf{y}} = \bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k}). \tag{30}$$

Eq. (30) is the exact version, valid for arbitrary data windows, of what Simons & Olhede (2013) implemented approximately and only for rectangular windows by discrete convolution of the theoretical spectral density $\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k})$ with the Fejér kernel $|D_K(\mathbf{k})|^2$, as in eq. (24). In that special case of a unitary window function, that is, for complete observations on a rectangular grid, eq. (29) evaluates to the triangular

$$W(\mathbf{y}) = (M-|m'|)(N-|n'|). \tag{31}$$

Generating Fourier coefficients $H(\mathbf{k})$ from the square root of the *blurred* spectral density $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$ of eq. (30), as Simons & Olhede (2013) did as a basis for simulation, ignores wavenumber *correlation* effects. In that case samples need to be generated on a spatial grid larger than needed, retaining only a central portion for analysis, to avoid wrap-around correlations or periodic realizations. Constructing spatial patches via eq. (22) on the space grid (19), the covariance of the results, $\langle\mathcal{H}(\mathbf{x})\mathcal{H}(\mathbf{x}')\rangle$, may be understood as a discrete approximation of the integral in eq. (4), with the spacings defined in eq. (21). Hermitian symmetry guarantees that the simulated fields are real, and their covariance

$$\langle\mathcal{H}(\mathbf{x})\mathcal{H}(\mathbf{x}')\rangle = \mathcal{C}_{\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}'), \tag{32}$$

is stationary. Furthermore, for large sample sets, $\sum_{\mathbf{k}}\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k}) \approx \sum_{\mathbf{k}}\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k})$, which establishes the desired correspondence

$$\text{var}\{\mathcal{H}(\mathbf{x})\} \approx \frac{(2\pi)^2\sum_{\mathbf{k}}\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k})}{(MN\Delta x\Delta y)} \approx \sigma^2. \tag{33}$$

Eq. (30) shows that the *expected periodogram* of the data, $\text{var}\{H(\mathbf{k})\} = \langle|H(\mathbf{k})|^2\rangle$, can be obtained via Fourier transformation of the autocovariance sequence of the sampling window. Fig. 2 shows this equivalence. From a sequence of simulations, we show, in Fig. 2*(a–c)*, one spatial-domain field $\mathcal{H}(\mathbf{x})$, its periodogram $|H(\mathbf{k})|^2$ on the corresponding normalized Fourier grid, and the expected periodogram, the blurred spectrum $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$, for the parameter set $\boldsymbol{\theta}$ shown at the top, $\sigma^2 = 1$ (in arbitrary field units), $\nu = 2.5$ and $\rho = 1$ (in units of the spatial grid spacing). In Fig. 2*(d–f)*, we show a unit-normalized square window cosine-tapered around the edges, the ratio of the average periodogram to its expectation, $\text{mean}\{|H(\mathbf{k})|^2\}/\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$, with its sample mean m and standard deviation s, and the average of the periodograms across 100 realizations, the sample variance $\text{var}\{H(\mathbf{k})\}$, which approximates the blurred spectrum $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$ shown directly above.

2.2.3 Simulation

Eq. (23) shows that the covariance of a finite set of gridded Fourier coefficients suffers both from blurring by the sampling kernel, as we have just illustrated and calculated explicitly, but also from correlation between the wavenumbers. To prepare for what is coming, we note, first, that if the off-diagonal terms in $\langle H(\mathbf{k})H^*(\mathbf{k}')\rangle$ decay rapidly enough, one could ignore them as the basis for simulations, taking $\langle H(\mathbf{k})H^*(\mathbf{k})\rangle \approx \bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$ as a point of departure, whether calculated on the interior domain of a doubled grid, using grid refinement, discrete convolution, and subsampling to approximate $\bar{\mathcal{S}}_{\boldsymbol{\theta}}$ (as was done by Simons & Olhede 2013), or exactly, via eq. (30). Here we do take wavenumber correlations into account for data simulation by switching to space-domain methods that use the *spatial* covariance, eq. (15), as

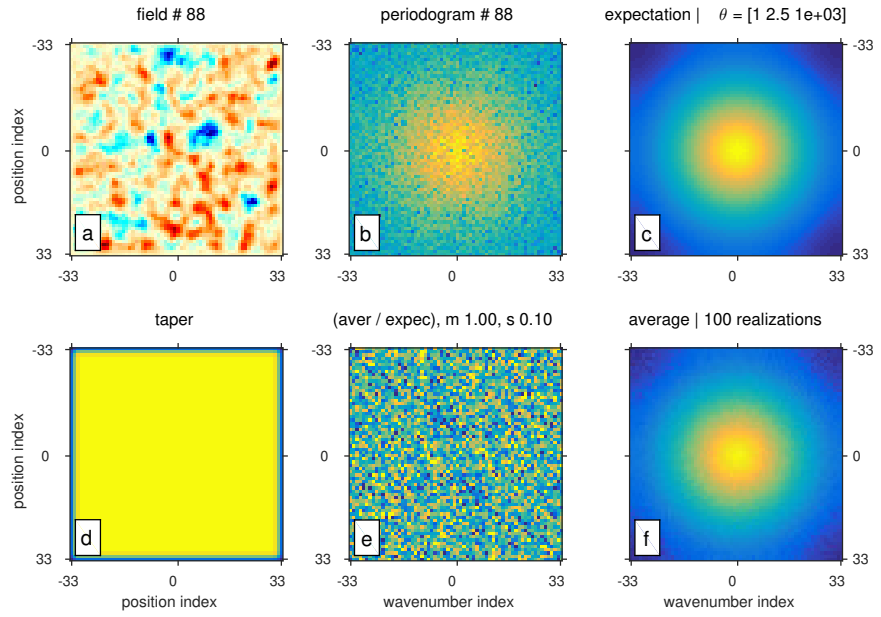

Figure 2. The Fejér blurred spectral density $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$ approximates the expectation of the periodogram $|H(\mathbf{k})|^2$, of gridded and cosine-tapered data generated from the population density $\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k})$. (*a*) A single realization $\mathcal{H}(\mathbf{x})$, (*b*) its modified periodogram $|H(\mathbf{k})|^2$, and (*c*) the blurred spectrum $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$. (*d*) The cosine-tapered unit window, (*e*) the ratio of the average periodogram to the blurred spectral density, and (*f*) the average periodogram, over 100 realizations.

their point of departure, via circulant embedding of the covariance matrix (Kroese & Botev 2015). Second, we will show empirically that we are able to ignore them when designing the debiased-Whittle likelihood (Guillaumin et al. 2017; Sykulski et al. 2019) to perform parameter estimates from sampled data, which is a central feature of this paper and its predecessors. Finally, we show that they will, however, play an important role in the calculation of the estimation variance of the maximum-likelihood estimates, using the results obtained by Guillaumin et al. (2022), and we discuss various algorithms to conduct the relevant calculations. This last fact stands in apparent contradiction to the material discussed by Simons & Olhede (2013), their Sections A6 and A8, which, in retrospect, have proven to be overly optimistic, asymptotically.

3 WHITHER WHITTLE ?

In this paper we develop a maximum-likelihood procedure that takes gridded input 'topographies' and estimates the three-element sets $\boldsymbol{\theta}$, see eq. (13), that contain the parameters of the isotropic Matérn spectral densities $\mathcal{S}_{\boldsymbol{\theta}}(k)$ or spatial covariances $\mathcal{C}_{\boldsymbol{\theta}}(r)$ by which we aim to sufficiently describe such planetary data patches. Before proceeding, we take a brief detour to illustrate, for the example of the variance, σ^2, *why* we embark on this journey. Additional motivation and considerations are offered by, among others, Vanmarcke (1983) and Stein (1999).

The variance σ^2 of a demeaned sample patch is *not* well estimated by its area-averaged sum of squares, which would amount to

$$s^2 = \frac{1}{MN}\sum_{\mathbf{x}} \mathcal{H}^2(\mathbf{x}) - \frac{1}{(MN)^2}\left(\sum_{\mathbf{x}} \mathcal{H}(\mathbf{x})\right)^2. \tag{34}$$

Indeed, the expectation of the 'sample variance' estimator, s^2, is biased by the co-variance between the samples, which itself is unknown. Using eqs (32) and (33), we find from eq. (34) that in expectation, approximately,

$$\langle s^2\rangle \approx \frac{1}{MN}\sum_{\mathbf{x}} \frac{(2\pi)^2\sum_{\mathbf{k}}\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k})}{(MN\Delta x\Delta y)} - \frac{1}{(MN)^2}\sum_{\mathbf{x}}\sum_{\mathbf{x}'}\mathcal{C}_{\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}') \approx \sigma^2 - \frac{1}{(MN)^2}\sum_{\mathbf{x}}\sum_{\mathbf{x}'}\mathcal{C}_{\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}') \tag{35}$$

$$\approx \sigma^2 - \frac{(2\pi)^2\bar{\mathcal{S}}_{\boldsymbol{\theta}}(0)}{MN\Delta x\Delta y}. \tag{36}$$

Eq. (35) applies quite generally, to stationary processes with spectral density $\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k})$ or covariance function $\mathcal{C}_{\boldsymbol{\theta}}(\mathbf{x})$. The error made in reducing the last term in (35) to the second term in eq. (36) should be interpreted as arising from the discretization of eq. (7). The appearance of the *blurred* spectral density $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(0)$ is due to the finite-sample effects by which the spatial grid is *relatively* coarse, and too small to comprise the lags at which the structure is completely decorrelated. Only for, *effectively*, uncorrelated white noise, $\mathcal{C}_{\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}') = \sigma^2\delta(\mathbf{x},\mathbf{x}')$, does eq. (35) reduce to the independent and identically distributed bias expression (Bendat & Piersol 2000)

$$\langle s^2\rangle \approx \sigma^2\left(1-\frac{1}{MN}\right). \tag{37}$$

If the spatial grid is fine enough, that is, if the pixel sizes Δx and Δy are small enough relative to ρ, the scale length of the correlation, thus

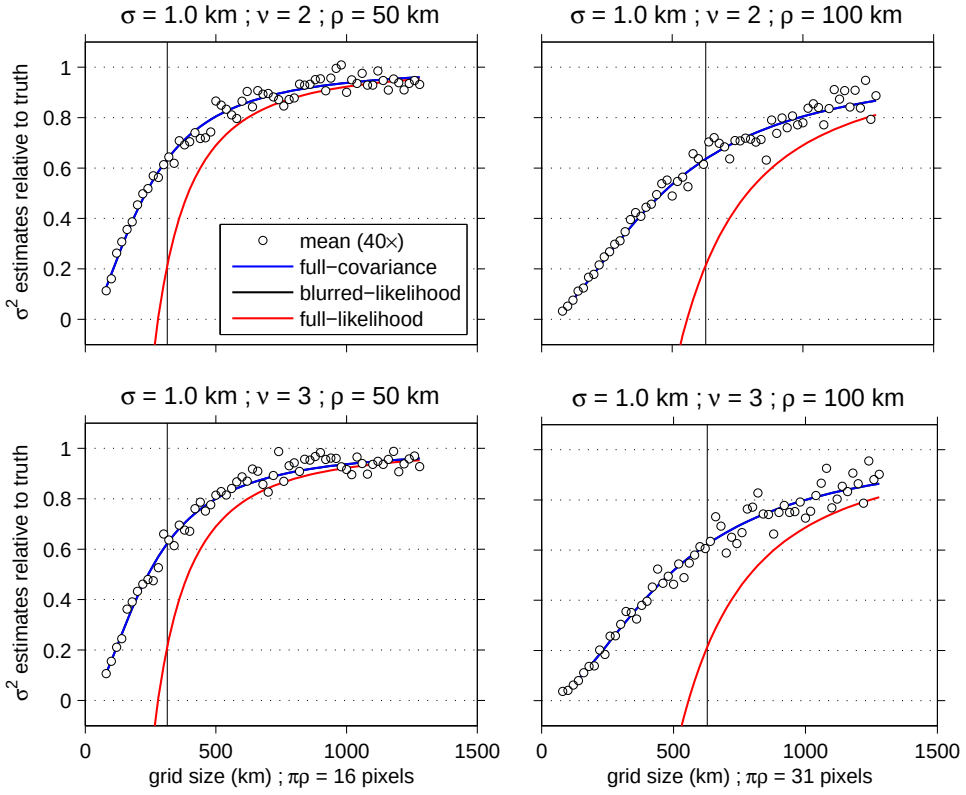

Figure 3. The sample variance s^2 systematically underestimates the true process variance σ^2. It is negatively biased by the presence of spatial correlation embodied by the Matérn parameters ν and ρ, listed in the titles. Black open circles ('mean') are averages of the sample variances s^2 for data patches of different sizes, as observed over 40 lattice simulations, normalized by the actual variance σ^2. Solid blue lines ('full-covariance') predict the average behavior by incorporating the bias according to eq. (35), evaluating eq. (15) on all of the pairwise distances available in the grids. Essentially hidden underneath the blue ones are solid black lines ('blurred-likelihood') resulting from calculations that use eq. (36). Solid red lines ('full-likelihood') are from eq. (38). As detailed in the text, the quality of the various approximations is to be interpreted in terms of the Matérn correlation parameters ν and ρ, in relation to the sampling spacings $(\Delta x, \Delta y)$, which were kept constant at 10 km, and the field sizes (M, N), which increased from left to right, as shown. The vertical black lines are drawn at the values $2\pi\rho$, a distance beyond which the bias in the sample variance estimator decreases to about a third of the true value, speaking empirically.

when the full behavior of the spatial covariance $\mathcal{C}_{\boldsymbol{\theta}}(\mathbf{x})$ is being accurately captured by the sampling density, $\mathcal{S}_{\boldsymbol{\theta}}(0)$ can again be substituted for $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(0)$ in eq. (36). In that case, using eq. (14) yields the form applicable to the isotropic Matérn density, namely

$$\langle s^2 \rangle \approx \sigma^2 \left(1 - \frac{\pi(\pi\rho)^2}{MN\Delta x \Delta y} \right). \tag{38}$$

While this last approximation is usually too crude for bias calculations, eq. (38) does explain the expected behavior that, the larger ρ, relative to the area of the study region, the more correlation will be present between the samples, causing us to underestimate the variance, and the more negatively biased the naive estimator eq. (34) will be. In real-world applications we will of course know neither the variance σ^2 nor the range ρ. Nor the smoothness ν, for that matter, knowledge of which might otherwise help us design better estimators, with ν held fixed.

Fig. 3 illustrates the arguments made so far in this section, for a variety of values of ν and ρ, as a function of grid size, and where the expectation of the estimate is formed by averaging over a great number of experiments. The naive variance estimator s^2 is biased, in a manner and for a reason that we understand intuitively, and are able to compute analytically. Rather than writing down expressions for the *variance* of the naive variance estimator s^2, we will illustrate its behavior on the basis of another suite of numerical experiments.

Fig. 4 (left panels) reveals that the estimation variance of s^2 is generally high (relatively speaking), and decaying too slowly (for our taste) with increasing grid size. In comparison, the quasi-maximum-likelihood estimator that we develop in the next section has properties that are far more favorable, as is readily, if proleptically, illustrated by the second suite of experiments shown in Fig. 4 (right panels, marked 'MLE'). Saving the details of its construction for the next section, inspecting the figure reveals that, as soon as the data patch size exceeds the correlation length of the Matérn process, the estimation variance of the maximum-likelihood variance estimator is acceptably low. Moreover, the estimation variance continues to decay at a pleasing rate, suggestive of its asymptotic unbiasedness.

While the examples thus far may have appeared anecdotal, it is hoped that they do convincingly hint at the agreeable qualities of the maximum-likelihood estimators, which we now discuss in more detail.

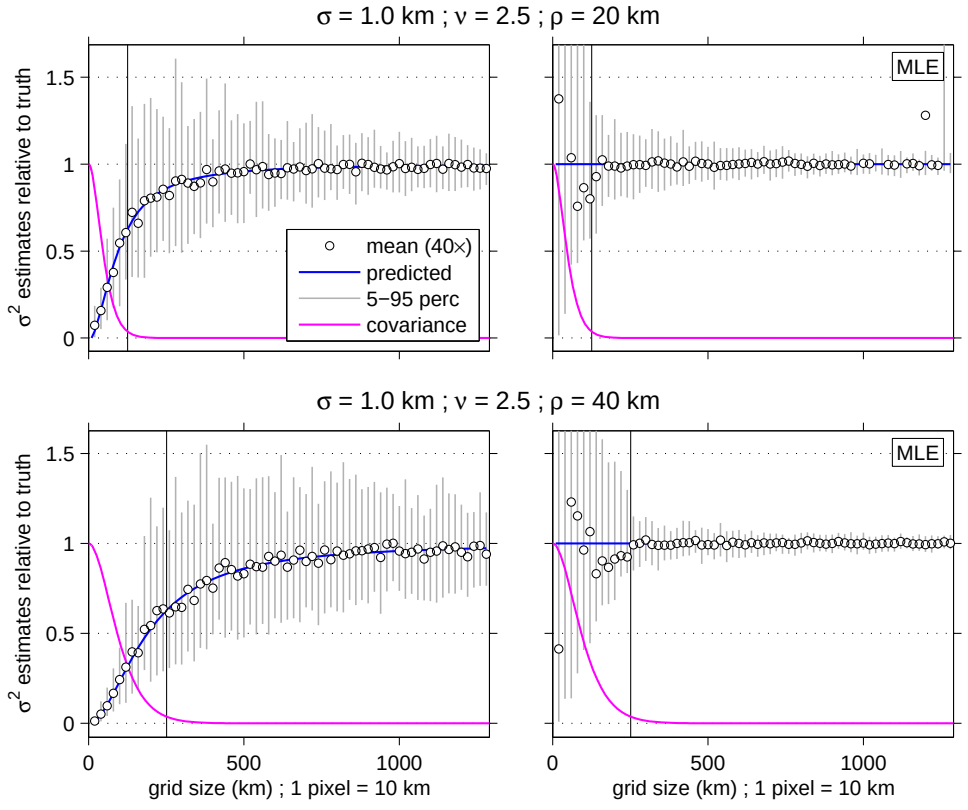

Figure 4. The sample variance s^2 is a biased, inconsistent, and inefficient estimator for the true process variance σ^2. The maximum-likelihood estimator is asymptotically unbiased, consistent and efficient. Conducting 40 lattice simulations on differently sized data patches, with Matérn parameters (σ^2, ν, ρ) as listed in the titles, the left panels show the behavior of the sample variance s^2, and the right panels that of the maximum-likelihood variance estimator ('MLE'), both normalized by the actual variance σ^2. The gray bars span the 5th to 95th percentiles of the estimates at the quoted patch sizes, the black open circles are the mean estimates, and the solid blue lines their predictions from eq. (35), as in Fig. 3. The magenta curves are the scaled spatial correlation functions, with the vertical black lines at $2\pi\rho$. The means of the MLE for field sizes smaller than $2\pi\rho$ were calculated over the 80th percentile of the estimates.

4 MAXIMUM-LIKELIHOOD THEORY

The material in this section is chiefly inspired by sections 4.3–4.8 and Appendix A6 of Simons & Olhede (2013). Cox & Hinkley (1974) remains an excellent reference, though more modern texts such as Pawitan (2001) and Severini (2001) are recommended. Our main device is the frequency-domain Whittle (1953, 1954) likelihood, modified to acknowledge edge effects by blurring the spectral density function.

4.1 Finite large-sample theory

Simons & Olhede (2013) introduced $\mathcal{L}(\boldsymbol{\theta})$, the likelihood for the Matérn parameters, which, as is at best acceptable only for large sample sizes, ignores the blurring in the isotropic spectral density $\mathcal{S}_{\boldsymbol{\theta}}(k)$ as well as the correlation induced between wavenumbers, in the form

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{MN}\sum_{\mathbf{k}}\left[\ln \mathcal{S}_{\boldsymbol{\theta}}(k) + \mathcal{S}_{\boldsymbol{\theta}}^{-1}(k)\,|H(\mathbf{k})|^2\right]. \tag{39}$$

Its first derivatives, with respect to each of the parameters $\theta \in \boldsymbol{\theta}$, are the components of the unblurred score vector $\boldsymbol{\gamma}(\boldsymbol{\theta})$, given by

$$\gamma_\theta(\boldsymbol{\theta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \theta} = -\frac{1}{MN}\sum_{\mathbf{k}} m_\theta(k)\left[1 - \mathcal{S}_{\boldsymbol{\theta}}^{-1}(k)\,|H(\mathbf{k})|^2\right], \tag{40}$$

with, easily obtained via differentiation of eq. (12), and listed as eq. (73) in the Appendix, the required expressions

$$m_\theta(k) = \mathcal{S}_{\boldsymbol{\theta}}^{-1}(k)\frac{\partial \mathcal{S}_{\boldsymbol{\theta}}(k)}{\partial \theta}. \tag{41}$$

Its second derivatives, with respect to two arbitrary parameters $\theta, \theta' \in \boldsymbol{\theta}$, are the components of the Hessian matrix $\mathbf{F}(\boldsymbol{\theta})$,

$$F_{\theta\theta'}(\boldsymbol{\theta}) = \frac{\partial \gamma_{\theta'}(\boldsymbol{\theta})}{\partial \theta} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial\theta\partial\theta'} = -\frac{1}{MN}\sum_{\mathbf{k}}\left[\frac{\partial m_{\theta'}(k)}{\partial\theta} + \left\{m_\theta(k)m_{\theta'}(k) - \frac{\partial m_{\theta'}(k)}{\partial\theta}\right\}\left\{\mathcal{S}_{\boldsymbol{\theta}}^{-1}(k)\,|H(\mathbf{k})|^2\right\}\right], \tag{42}$$

with the nonvanishing derivatives $\partial_\theta m_{\theta'}(k)$ given as eqs (74)–(75) in the Appendix.

The negative expectation of $\mathbf{F}(\boldsymbol{\theta})$ is the Fisher matrix $\boldsymbol{\mathcal{F}}(\boldsymbol{\theta})$, which does not depend on the data as $\langle \mathcal{S}_{\boldsymbol{\theta}}^{-1}(k)\,|H(\mathbf{k})|^2\rangle = 1$, and thus

$$\mathcal{F}_{\theta\theta'}(\boldsymbol{\theta}) = -\left\langle \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial\theta\partial\theta'} \right\rangle = \frac{1}{MN}\sum_{\mathbf{k}} m_\theta(k)\, m_{\theta'}(k). \tag{43}$$

The inverse of the Fisher matrix is the information matrix, $\boldsymbol{\mathcal{F}}^{-1} = \boldsymbol{\mathcal{J}}$. Denoting the true parameter set as $\boldsymbol{\theta}_0$, with elements θ_0, and the maximum-likelihood estimate as $\hat{\boldsymbol{\theta}}$, with elements $\hat{\theta}$, the presumed normality of the Fourier coefficients (25) implies the distribution

$$\sqrt{MN}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)) = \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{J}}(\boldsymbol{\theta}_0)), \tag{44}$$

from which we will seek to construct $100\times(1-\beta)$ per cent confidence intervals about the estimates, using the values $z_{\beta/2}$ at which the standard-normal distribution reaches a cumulative probability of $(1-\beta/2)$, as follows:

$$\hat{\theta} - z_{\beta/2}\frac{\mathcal{J}_{\theta\theta}^{1/2}(\hat{\boldsymbol{\theta}})}{\sqrt{MN}} \le \theta_0 \le \hat{\theta} + z_{\beta/2}\frac{\mathcal{J}_{\theta\theta}^{1/2}(\hat{\boldsymbol{\theta}})}{\sqrt{MN}}. \tag{45}$$

The relations in this section are theoretical quantities derived by Simons & Olhede (2013) that, strictly speaking, apply only to the 'population' case, on domains of infinite extent, and we note that caveats about consistency apply for MN large in the fixed-domain asymptotic regime, as we alluded to in Sec. 1. While even under increasing-domain asymptotics, there may be few rigorous results on maximum likelihood estimates for the Matérn model when the smoothness parameter is treated as unknown, the *(highly) significant correlation contribution* introduced by Guillaumin et al. (2022) helps illuminate the estimability of, and the identifiability between, the three parameters, and the separation of different processes under various spatial sampling mechanisms, which will enable pursuing these issues further. In the 'sample' case of discretized, windowed likelihood analysis, we follow Simons & Olhede (2013) and Guillaumin et al. (2022) in replacing the Matérn spectral density $\mathcal{S}_{\boldsymbol{\theta}}(k)$ in eq. (39) with a suitably blurred version, $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(k)$, to acknowledge the effects of finite sampling. In what follows, we will explore the implications of sampling and bounding on the uncertainty estimates of the parameters.

When using the Whittle likelihood as defined in eq. (39) to produce estimators for time series, the behavior of the spectral estimates is very well understood, see for example Dzhaparidze & Yaglom (1983). Exactly how the Whittle likelihood is formulated may vary—in terms of an integral or a sum, with the log covariance term included or not (Hosoya & Taniguchi 1982), and with and without tapering (Dahlhaus 1984). In two dimensions and higher, however, it is recognized that the boundary effects become dominant, which has led to alternative formulations (Guyon 1982; Deb et al. 2017; Guillaumin et al. 2022). Grainger et al. (2025) treat the non-parametric asymptotics of (even non-separable) tapers coherently. Here we follow Guillaumin et al. (2022) in producing estimators from the forthcoming spectral eq. (46), which removes bias from boundary effects, can be tapered, maintains the statistical efficiency of a spatial likelihood, and remains computationally competitive. Should the parameters of the model become sample-size dependent, no existing theory is available yet.

4.2 Finite sampled data: heuristics

For sampled data, Simons & Olhede (2013) defined the likelihood of observing the data $\mathcal{H}(\mathbf{x})$ under the spectral model (12)–(13) parameterized by $\boldsymbol{\theta}$. This 'blurred likelihood' $\bar{\mathcal{L}}(\boldsymbol{\theta})$ instead of $\mathcal{L}(\boldsymbol{\theta})$, is given in terms of the Fourier coefficients of the gridded and windowed data, $H(\mathbf{k})$ in eq. (25), and of the blurred isotropic spectral density, $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$ of eq. (30), summed over all wavenumbers in the grid (21), by

$$\bar{\mathcal{L}}(\boldsymbol{\theta}) = -\frac{1}{MN}\sum_{\mathbf{k}}\left[\ln \bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k}) + \bar{\mathcal{S}}_{\boldsymbol{\theta}}^{-1}(\mathbf{k})\,|H(\mathbf{k})|^2\right]. \tag{46}$$

Compare eq. (46) with eq. (39): the only difference is the acknowledgment of the spectral blurring effect of sampled data. Eq. (46) is the quantity that we maximize, under positivity constraints, for the parameter vector $\boldsymbol{\theta}$, thereby defining the quasi-maximum-likelihood estimate $\hat{\boldsymbol{\theta}}$ to be obtained by solving for the 'score' vector $\bar{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ of numerical first derivatives of the blurred likelihood, in the sense

$$\bar{\boldsymbol{\gamma}}(\hat{\boldsymbol{\theta}}) = \mathbf{0}. \tag{47}$$

Satisfying eq. (47) to find $\hat{\boldsymbol{\theta}}$, for example by an iterative function minimization routine, requires repeated evaluation of the spectral density (12) on the grid (21), with the blurring implemented convolutionally (on a refined and subsequently reinterpolated grid) or else exactly, as in eq. (30). Without entering into more details for now, Fig. 5 shows the results of a suite of experiments conducted to illustrate the performance of our numerical method that recovers each of the three Matérn parameters (σ^2, ν, ρ), as a function of grid size (measured in terms of the correlation length $\pi\rho$). As the right-hand side panels of Fig. 4 showed for the variance estimate, the maximum-likelihood estimates are very well-behaved, from about the point where the grid size reaches a linear dimension of about $\pi\rho$. The procedure almost surely yields low-variance and practically unbiased estimates from a grid size of about $2\pi\rho$ onward, as Fig. 5 shows empirically.

The mathematical form and geometry of the blurred likelihood function are what, fundamentally, controls the observed behaviors. Simons & Olhede (2013) only considered convolutional approaches to blurring, and were limited in their ability to acknowledge the spectral interaction induced by the applied data windows on the parameter covariance estimates. The first and second derivatives of the blurred likelihood are *not* simply obtained by substituting $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(k)$ for $\mathcal{S}_{\boldsymbol{\theta}}(k)$ in eqs (40) and (42), since the factors $m_\theta(k)$ that appear in the expressions for the score and the Hessian are *analytical* derivatives that involve the *unblurred* spectral density $\mathcal{S}_{\boldsymbol{\theta}}(k)$. Replacing $\mathcal{S}_{\boldsymbol{\theta}}(k)$ by $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(k)$ in eq. (40) yields *reasonable* approximations for the likelihood derivatives, which compare *relatively* favorably to their numerical counterparts—especially for large sample sizes. Most numerical optimization routines will be able to maximize eq. (46), solving eq. (47) without being given

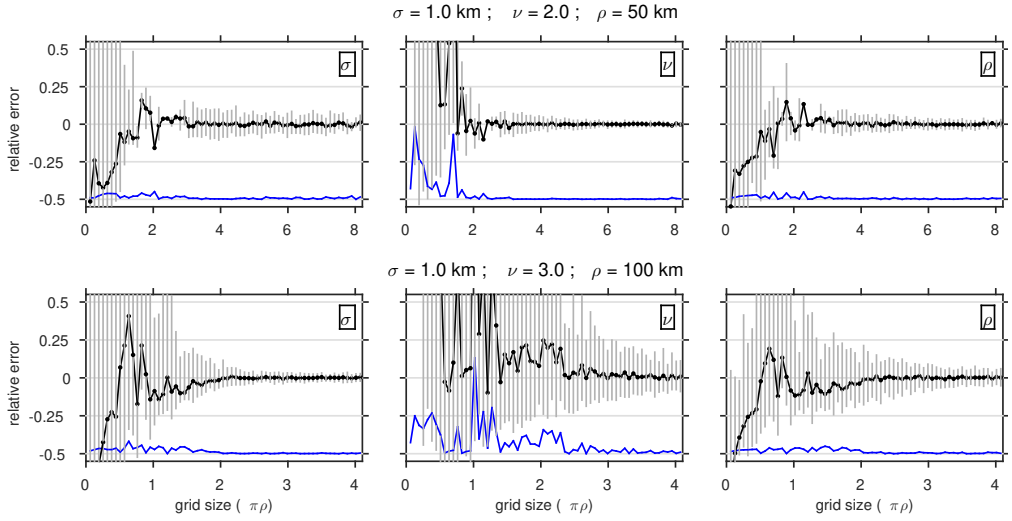

Figure 5. Behavior of the relative error of the Whittle maximum-likelihood estimators of the Matérn parameters (σ^2, ν, ρ) for the two sets of true values listed in the titles, conducted on square lattices growing in size from $M = N = 2$ to $M = N = 128$ pixels of size $\Delta x = \Delta y = 10$ km, quoting $M\Delta x$ in multiples of $\pi\rho$. Gray bars cover the 5th through 95th percentiles of the estimates for each set of up to 40 simulations. Black filled circles are the means of the estimates, computed over the 80th percentile of the sets for fields whose linear dimension $M\Delta x < 2\pi\rho$, but over the full set of up to 40 estimates beyond that size. With growing grid size, the estimates reveal themselves to be less biased with shrinking variance. The blue line tracks the absolute value of the relative error multiplied by $\sqrt{MN}$, offset by -0.5 for visual clarity.

explicit expressions for the score and the Hessian. However, in order to derive accurate confidence intervals on our estimated parameters, we do need access to the expected value of the second-order derivatives of the *actual* likelihood that is being maximized.

Numerical experiments and theoretical considerations along the lines suggested in their Appendix A8 tempted Simons & Olhede (2013) into concluding that eq. (45) could be used to construct confidence intervals for the solutions of eq. (47) in our present case of univariate two-dimensional analysis. Under the viewpoint espoused in their eqs (A84)–(A87), the blurred spectrum is an *additive* correction term (small for smoothly varying spectra) away from the original. In this framework, neglecting to blur the Fisher matrix (not to mention neglecting wavenumber correlations) was believed to have an altogether negligible effect on the estimation variances based on its inverse, even if blurring the likelihood is absolutely essential to arrive at the estimate in the first place. However, the ability of the *unblurred* Fisher matrix to help predict the variance of the parameters derived via maximization of the *blurred* likelihoods turns out to be poor, especially as concerns the variance and correlation parameters σ^2 and ρ. The unblurred expression, eq. (43), of the Fisher matrix provides an asymptotic but ultimately inadequate match to the average of the numerical Hessian for real-world sampling scenarios.

Not accounting for wavenumber correlation proved to be another stumbling block. Simons & Olhede (2013) conceived of approximations to account for wavenumber correlation involving a *multiplicative* correction term (their eqs A56–A58). For very large sample sizes this correction term approaches unity. Contrary to the optimism they expressed, uncertainty estimates for the maximizers of eq. (46) that rely on eq. (45) are inadequate for all but the largest sample sizes. A heuristic way of determining the estimation variance for the recovered parameters when actual data are being investigated is to generate synthetics with features identical to those of the gridded data, from models with Matérn parameters given by previously obtained solutions, then estimating their parameters a number of times, and learning from their distribution what the likely uncertainty ranges for the parameters of the actual data patches might be, as in Fig. 5. However justified, little transferable knowledge is gained in the process, and the procedure is cumbersome and time-consuming.

Guillaumin et al. (2022) showed the way forward by further developing the theory of likelihood analysis for finite sampled data, on which we rely to develop the practical methods offered in the next sections. They include the ability to calculate uncertainty estimates on the parameters from first principles. The next section provides a complete description of the entire workflow.

4.3 Finite sampled data: full theory

For sampled data, the likelihood involves $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$, the blurred spectral density (eqs 24 and 30), and the modified periodogram of the data $\mathcal{H}(\mathbf{x})$,

$$\bar{\mathcal{L}}(\boldsymbol{\theta}) = -\frac{1}{MN}\sum_{\mathbf{k}}\left[\ln \bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k}) + \bar{\mathcal{S}}_{\boldsymbol{\theta}}^{-1}(\mathbf{k})\,|H(\mathbf{k})|^2\right], \tag{48}$$

where $H(\mathbf{k})$ is the windowed Fourier transform of the data $\mathcal{H}(\mathbf{x})$, for an arbitrary unit-normalized window $w(\mathbf{x})$, repeating eq. (25),

$$H(\mathbf{k}) \equiv \frac{1}{2\pi}\left(\frac{\Delta x \Delta y}{MN}\right)^{\frac{1}{2}} \sum_{\mathbf{x}} w(\mathbf{x})\mathcal{H}(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{49}$$

and the Matérn spectral density $\mathcal{S}_{\boldsymbol{\theta}}$ whose parameters we aim to recover (see eq. 12) is exactly blurred (hence no longer isotropic, see Fig. 2) to account for finite-sample effects via the intermediary of the isotropic spatial Matérn covariance $\mathcal{C}_{\boldsymbol{\theta}}(y)$, see eq. (15), repeating eq. (30),

$$\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k}) = \frac{1}{(2\pi)^2}\left(\frac{\Delta x \Delta y}{MN}\right)\sum_{\mathbf{y}} W(\mathbf{y})\mathcal{C}_{\boldsymbol{\theta}}(y)e^{-i\mathbf{k}\cdot\mathbf{y}} = \operatorname{var}\{H(\mathbf{k})\}, \tag{50}$$

and with W the autocorrelation of the sampling window, obtainable via FFT, that is, repeating eq. (29) over the relevant grid (lags),

$$W(\mathbf{y}) = \sum_{\mathbf{x}} w(\mathbf{x})w(\mathbf{x}+\mathbf{y}). \tag{51}$$

The solution $\hat{\boldsymbol{\theta}}$ is found by maximization of eq. (48), requiring the vanishing of the score $\boldsymbol{\nabla}_{\boldsymbol{\theta}}\bar{\mathcal{L}}(\boldsymbol{\theta}) = \bar{\boldsymbol{\gamma}}(\boldsymbol{\theta})$, whose components are given by

$$\bar{\gamma}_\theta(\boldsymbol{\theta}) = -\frac{1}{MN}\sum_{\mathbf{k}} \bar{m}_\theta(\mathbf{k})\left[1 - \bar{\mathcal{S}}_{\boldsymbol{\theta}}^{-1}(\mathbf{k})\,|H(\mathbf{k})|^2\right], \tag{52}$$

with the blurred equivalents to eq. (41) again obtained exactly via the intermediary of eq. (50) as

$$\bar{m}_\theta(\mathbf{k}) = \bar{\mathcal{S}}_{\boldsymbol{\theta}}^{-1}(\mathbf{k})\frac{\partial \bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})}{\partial\theta} = \frac{\bar{\mathcal{S}}_{\boldsymbol{\theta}}^{-1}(\mathbf{k})}{(2\pi)^2}\left(\frac{\Delta x\Delta y}{MN}\right)\sum_{\mathbf{y}} W(\mathbf{y})\,\frac{\partial \mathcal{C}_{\boldsymbol{\theta}}(y)}{\partial\theta}e^{-i\mathbf{k}\cdot\mathbf{y}}. \tag{53}$$

The requisite derivatives of the spatial covariance $\partial_\theta \mathcal{C}_{\boldsymbol{\theta}}$ are obtained via differentiation of eq. (15) and listed as eqs (76)–(78) in the Appendix.

The components of the Fisher matrix $\bar{\boldsymbol{\mathcal{F}}}(\boldsymbol{\theta})$, with respect to two arbitrary Matérn parameters $\theta, \theta' \in \boldsymbol{\theta}$, are now given by

$$\bar{\mathcal{F}}_{\theta\theta'}(\boldsymbol{\theta}) = \frac{1}{MN}\sum_{\mathbf{k}} \bar{m}_\theta(\mathbf{k})\,\bar{m}_{\theta'}(\mathbf{k}). \tag{54}$$

As Simons & Olhede (2013) (their eq. 138), but now following Guillaumin et al. (2022) (their eq. 36), the parameter estimation variance,

$$\operatorname{cov}(\hat{\boldsymbol{\theta}}) \approx \bar{\boldsymbol{\mathcal{F}}}^{-1}(\boldsymbol{\theta}_0)\operatorname{cov}\{\bar{\boldsymbol{\gamma}}(\boldsymbol{\theta}_0)\}\,\bar{\boldsymbol{\mathcal{F}}}^{-1}(\boldsymbol{\theta}_0), \tag{55}$$

where in practice we substitute $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta_0}$, to which the estimator converges in probability, requires the additional calculation of the covariance of the score *without* neglecting the correlation between wavenumbers,

$$\operatorname{cov}\{\bar{\gamma}_\theta, \bar{\gamma}_{\theta'}\} = \frac{1}{(MN)^2}\sum_{\mathbf{k}}\sum_{\mathbf{k}'} \bar{m}_\theta(\mathbf{k})\,\frac{\operatorname{cov}\{|H(\mathbf{k})|^2, |H(\mathbf{k}')|^2\}}{\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k}')}\,\bar{m}_\theta(\mathbf{k}'), \tag{56}$$

see Guillaumin et al. (2022) (their eq. 37), and compare Simons & Olhede (2013) (their eq. A54, which should have quoted blurred quantities). Eq. (56) implies that we require the covariance of the windowed periodogram $|H(\mathbf{k})|^2$, which under standard theory using the Isserlis (1918) theorem for Gaussian processes is (Percival & Walden 1993; Stein 1995; Simons & Olhede 2013, their eq. A57)

$$\operatorname{cov}\{|H(\mathbf{k})|^2, |H(\mathbf{k}')|^2\} = \left|\operatorname{cov}\{H(\mathbf{k}), H^*(\mathbf{k}')\}\right|^2 + \left|\operatorname{cov}\{H(\mathbf{k}), H(\mathbf{k}')\}\right|^2. \tag{57}$$

4.4 Covariance of the estimates

To obtain $\operatorname{cov}(\hat{\boldsymbol{\theta}})$ in eq. (55), $\operatorname{cov}\{\bar{\boldsymbol{\gamma}}_{\boldsymbol{\theta}}\}$ in eq. (56) can be stochastically approximated by repeated simulation of the random field $\mathcal{H}(\mathbf{x})$, see Sec. 2.2.3, and calculating $\bar{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ of eq. (52) via numerical differentiation. Used inside eq. (55), this readily produces acceptable approximations for the parameter covariance in all cases we considered. Eq. (56) can also be calculated *exactly* using one of two methods.

The first is by direct calculation of the covariance between the windowed periodograms in eq. (57), rewritten (Walden et al. 1994) as

$$\begin{aligned}
\operatorname{cov}\{|H(\mathbf{k})|^2, |H(\mathbf{k}')|^2\} &= \left|\langle H(\mathbf{k})H(\mathbf{k}')\rangle\right|^2 + \left|\langle H(\mathbf{k})H^*(\mathbf{k}')\rangle\right|^2 &(58)\\
&= \left|\mathsf{Q}\,\langle \mathcal{H}(\mathbf{x})\mathcal{H}(\mathbf{x}')\rangle\,\mathsf{Q}'\right|^2 + \left|\mathsf{Q}\,\langle \mathcal{H}(\mathbf{x})\mathcal{H}(\mathbf{x}')\rangle\,\mathsf{Q}^{*\prime}\right|^2 &(59)\\
&= \left|\mathsf{Q}\,\mathcal{C}_{\boldsymbol{\theta}}\,\mathsf{Q}'\right|^2 + \left|\mathsf{Q}\,\mathcal{C}_{\boldsymbol{\theta}}\,\mathsf{Q}^{*\prime}\right|^2, &(60)
\end{aligned}$$

where, abusing notation ever so slightly, Q is the discrete-Fourier transform matrix operator that transforms windowed spatial-domain observations to the spectral domain, Q' and $\mathsf{Q}^{*\prime}$ its transpose and conjugate transposes, respectively, and $\mathcal{C}_{\boldsymbol{\theta}}$ the Matérn autocovariance sequence at all available lags. Fig. 6 illustrates this computational procedure for a spatial observation grid with a unit-window taper. The top row shows the covariance sequence $\mathcal{C}_{\boldsymbol{\theta}}(\|\mathbf{x}-\mathbf{x}'\|)$ with all two-dimensional lag-distances unwrapped along the two axes labeled $\mathbf{x}(:)$ and $\mathbf{x}'(:)$, followed by the covariance of the periodograms as the sum of the individual (pseudo-)covariance (Neeser & Massey 1993) terms as they appear in eq. (57). The bottom row shows the $\mathbf{k} = \mathbf{k}'$ diagonal elements of each of the latter, wrapped as k_x and k_y about the zero wavenumber in the center, following the independently (via eq. 50) calculated square of the blurred spectral density, as labeled. The main diagonal of the periodogram covariance, its variance $\operatorname{cov}\{|H(\mathbf{k})|^2, |H(\mathbf{k})|^2\} = \operatorname{var}\{|H(\mathbf{k})|^2\}$, matches the square of the blurred spectral density $\bar{\mathcal{S}}^2_{\boldsymbol{\theta}}(\mathbf{k})$.

The second implementation for calculating the covariance of the blurred and correlated score is by considering the terms in eq. (58). Returning to the integral form of eq. (23), rewritten with the explicit acknowledgment of a window function $w(\mathbf{x})$, which may not be unitary, used as a subscript, we write, integrating over the Nyquist plane of frequencies, for the case of the second term of eq. (58),

$$\langle H(\mathbf{k})H^*(\mathbf{k}')\rangle = \iint_{\mathbf{k}} D_K(\mathbf{k}-\mathbf{k}'')D_K^*(\mathbf{k}'-\mathbf{k}'')\,\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k}'')\,d\mathbf{k}'' \quad \text{where} \quad D_K(\mathbf{k}) = \frac{1}{2\pi}\left(\frac{\Delta x\Delta y}{MN}\right)^{\frac{1}{2}}\sum_{\mathbf{x}} w(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}. \tag{61}$$

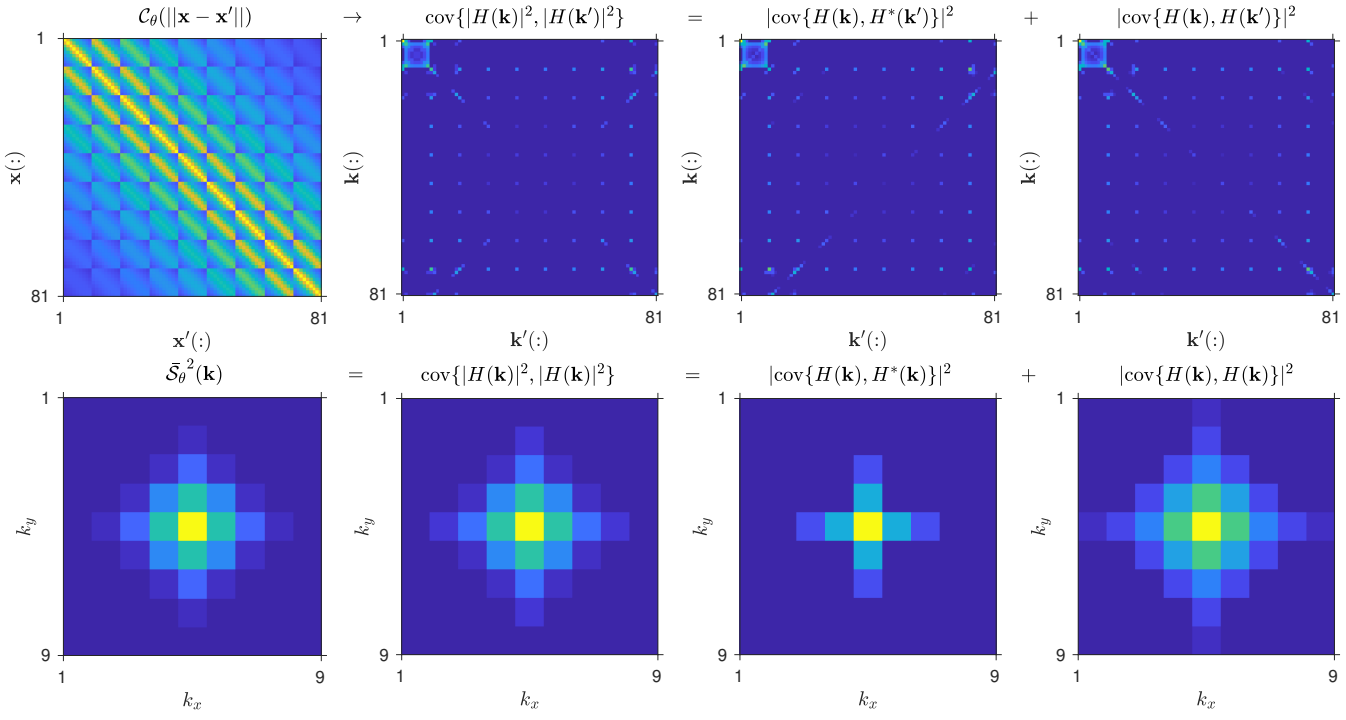

Figure 6. Sampling and windowing effects on the covariance between periodograms: blurring and correlation illustrated for a Matérn model with parameters $\boldsymbol{\theta}_0 = (\sigma^2, \nu, \rho) = (1, 0.5, 2)$ on a grid of $M = N = 9$ pixels of size $\Delta x = \Delta y = 1$ in arbitrary units. (*Top row*) The spatial covariance $\mathcal{C}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')$ on the grid that gives rise to the periodogram covariance $\mathrm{cov}\{|H(\mathbf{k})|^2, |H(\mathbf{k}')|^2\}$, which is the sum of two (pseudo-)covariance terms, see eq. (57). The off-diagonal terms express the correlation between periodograms at different frequencies; the diagonal terms embody their blurring. (*Bottom row*) The squared blurred spectral density $\bar{\mathcal{S}}^2_{\boldsymbol{\theta}}(\mathbf{k})$ is close to the variance of the periodograms $\mathrm{var}\{|H(\mathbf{k})|^2\}$, as shown in the last three panels, which are identical to the diagonals of the panels above them, wrapped into a two-dimensional spectral grid.

At a fixed distance between wave vectors $\Delta\mathbf{k} = \mathbf{k} - \mathbf{k}'$, eq. (61) amounts to a simple convolution of the theoretical spectral density with a kernel that is the product of two relatively offset Fourier series, $D_K(\mathbf{k})D_K^*(\mathbf{k} + \Delta\mathbf{k})$, which we can in turn calculate as a Fourier series via the convolution theorem. When $\Delta\mathbf{k} = \mathbf{0}$, without offset, we recover the variance of the windowed Fourier coefficients as the expectation of the periodogram via convolution of the spectral density with the modified Fejér kernel $|D_K(\mathbf{k})|^2$. Analogously with eq. (50) we compute

$$\langle H(\mathbf{k})H^*(\mathbf{k}')\rangle = \frac{1}{(2\pi)^2}\left(\frac{\Delta x \Delta y}{MN}\right)\sum_{\mathbf{y}} W(\mathbf{y}, \Delta\mathbf{k})\, \mathcal{C}_{\boldsymbol{\theta}}(y)\, e^{-i\mathbf{k}\cdot\mathbf{y}}, \quad \text{with} \quad \Delta\mathbf{k} = \mathbf{k} - \mathbf{k}', \tag{62}$$

adapting the autocorrelation of the energy-normalized analysis window as follows:

$$W(\mathbf{y}, \Delta\mathbf{k}) = \sum_{\mathbf{x}} w(\mathbf{x})w(\mathbf{x}+\mathbf{y})\, e^{i\Delta\mathbf{k}\cdot(\mathbf{x}+\mathbf{y})}. \tag{63}$$

Proceeding similarly, after reindexing and taking into account relevant symmetries for the pseudocovariance first term in eq. (58), yields

$$\langle H(\mathbf{k})H(\mathbf{k}')\rangle = \iint_{\mathbf{k}} D_K(\mathbf{k} - \mathbf{k}'')D_K(\mathbf{k}' + \mathbf{k}'')\, \mathcal{S}_{\boldsymbol{\theta}}(\mathbf{k}'')\, d\mathbf{k}''. \tag{64}$$

This procedure requires one Fast Fourier Transform per diagonal increasingly offset from the main diagonal in $(\mathbf{k}, \mathbf{k}')$ space, further multiplied with $\bar{m}_\theta(\mathbf{k})/\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$ and $\bar{m}_\theta(\mathbf{k}')/\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k}')$ and normalized into eq. (56). Like the procedure in eqs (58)–(60), it is exact (as opposed to Riemann discretization of eqs (61) and (64), as originally proposed by Guillaumin et al. 2022).

Assembling all the various pieces, we now have three different methods to obtain the desired estimation variance in eq. (56): likelihood gradient sampling, direct calculation via the DFT matrix, and per-diagonal Fourier transformation of the spatial covariance. The first is fast and has shown good behavior in all of our tests, see Sec. 4.5. The second and third produce identical results, with the matrix implementation being memory intensive, and the per-diagonal sequential calculation still computationally heavy as it calls on MN FFTs, each of which requiring $MN\log(MN)$ operations. As shown in Fig. 6, not all interaction terms are large, which has inspired randomized approaches to capture their contributions (Stein et al. 2004), e.g., via Markov-Chain Monte Carlo (Metropolis & Ulam 1949; Hastings 1970) or importance sampling. Managing the interplay of the stochastic process being sampled and the characteristics of the sampling strategy is amenable to formal treatment via the *significant correlation contribution* as developed by Guillaumin et al. (2017, 2022).

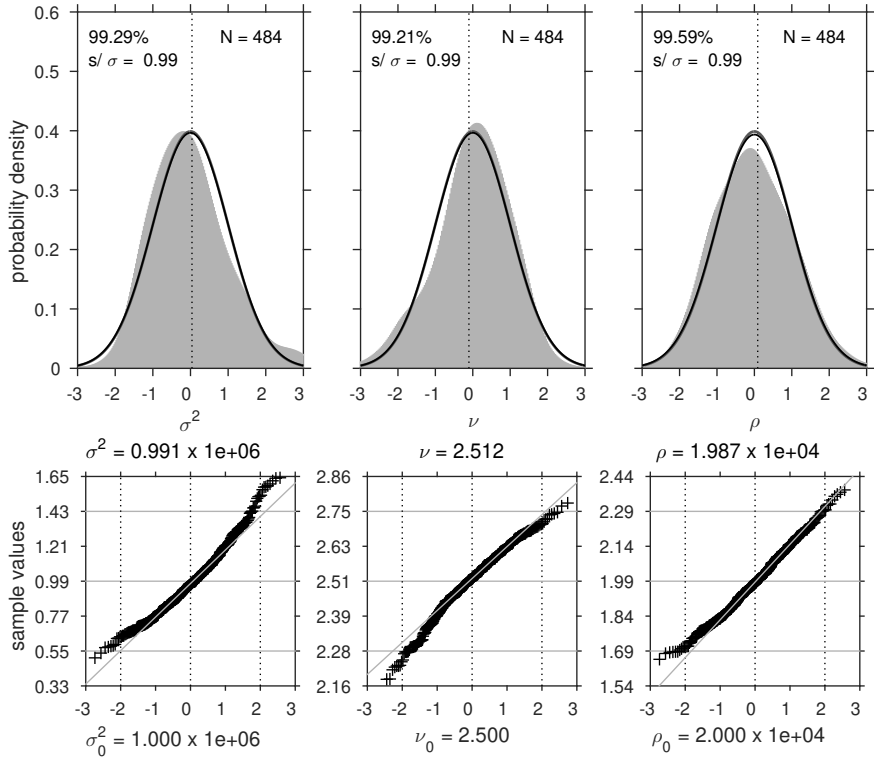

Figure 7. Matérn parameter maximum-likelihood estimation statistics for 484 spatial-covariance embedding simulations carried out on a 101×111 grid, with spacings $\Delta x = \Delta y = 10$ km, and true values of $\sigma_0^2 = 1$ km^2, $\nu_0 = 2.5$, $\rho_0 = 20$ km, recovered via maximization of the exactly blurred uncorrelated likelihood (48). The estimates average to $\sigma^2 = 0.91 \pm 0.22$ km^2, $\nu = 2.51 \pm 0.12$, $\rho = 19.87 \pm 1.509$ km, quoting one observed standard deviation. In the top row, the thick dark gray line overlying the light gray shaded smooth kernel-density estimate of the ensemble of simulation and estimation outcomes is the Gaussian derived from the covariance of the ensemble. The (nearly entirely overlapping) thick black line centered on the mean estimate is based on the covariance exactly calculated from eq. (55) evaluated at the truth, indicated by the dotted vertical line in the top panels. In the bottom row, quantile-quantile plots compare the empirical distribution to a normal distribution standardized using the mean and standard deviation of the ensemble.

4.5 Synthetic numerical examples

While all of our results discussed thus far hold even when the analysis window $w(\mathbf{x})$ is an irregularly bounded and possibly incompletely sampled subset of the regular grid (19), in the remainder of this paper we focus on completely sampled $M \times N$ rectangular grids.

Fig. 7 is a representative illustration of the behavior of the estimator $\hat{\boldsymbol{\theta}}$ of eqs (46)–(48), in each of the Matérn parameters σ^2, ν, and ρ. We conducted five hundred inversions (484 of which successfully yielded estimates) on independently generated simulations (via circulant embedding of the spatial covariance), and studied the empirical distribution of the estimates, compared to the theoretical expression of eq. (55) using the exact method of eqs (58)–(60). Invariably, the estimates are nearly unbiased, and nearly universally Gaussian distributed, as can be seen from the histograms and the quantile-quantile plots. The top row shows the smoothly estimated (Botev et al. 2010) standardized probability density functions (shaded) of the values recovered in this experiment of sample size 101×111. The abscissas were truncated to within ±3 of the empirical standard deviation about the origin at the mean estimate; the percentage of the values captured by this truncation is listed in the top left of each graph. The thick gray line summarizes the empirical distributions as a Gaussian with the observed mean and variance in this set of simulations. The thick black line uses the corresponding theoretical values calculated from eq. (55). The ratio of the empirical to the theoretical standard deviation is shown listed as s/σ for each of the parameters. The dotted black line marks the true parameter values. The bottom row shows the quantile-quantile plots of the theoretical (abscissa, horizontal) versus the empirical (ordinate, vertical) distributions. The averages of the recovered values σ^2, ν, and ρ are listed at the top of the second row of graphs. The true parameter values σ_0^2, ν_0, and ρ_0 are listed at the bottom. We attribute the slight asymmetry of the histograms and the small bends at the edges of the quantile-quantile plot to small-sample effects on minimum/maximum order statistics, and to all Matérn parameters being necessarily positive.

Fig. 8 is a rendering of the likelihood (48) that is being navigated towards an estimate. We purposely picked a data set from the series of five hundred runs that yielded an estimate very close to the truth in order to clean up the axis labeling, but the likelihood surfaces are typically very similar. Every panel occupies ±3 of the empirical standard deviations around the estimate. We found ten parameter value pairs evenly spaced out to $2\times$ the empirical standard deviations from the estimate, and calculated their likelihoods to use as the shaded contours.

Fig. 9 summarizes the same experiment as cross plots of the recovered parameters with their empirical summaries (Aster et al. 2018) as the ellipsoidal 68% confidence region (thick gray line), and showing the equivalent error ellipses from the theoretically predicted covariances (thick black line), which match the observations closely, albeit not perfectly. Every panel shows all pairs of parameter estimates (gray circles),

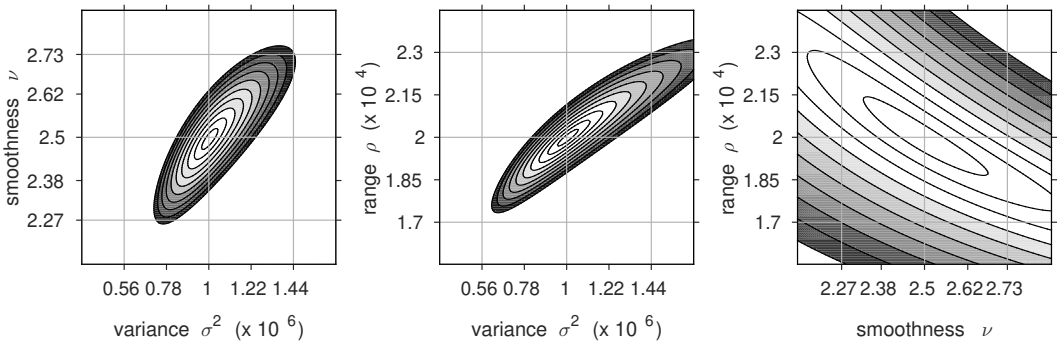

Figure 8. An example of a likelihood (eq. 48) surface for simulated data from the experiment reported in Fig. 7, shown for one estimate very close to the truth. Every two-dimensional panel is a slice at the value of the third estimated parameter. We draw and shade ten contours in equal increments, from the likelihood at the estimate in the center, to the likelihood of the parameter pair two observed standard deviations out, in the positive quadrant of the ensemble.

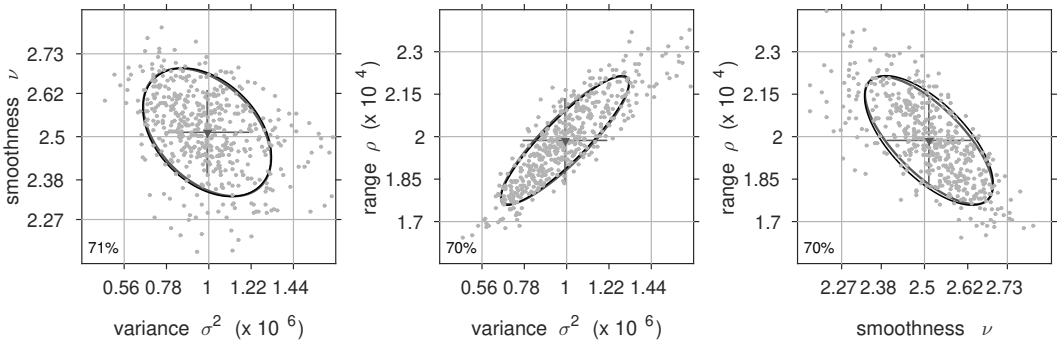

Figure 9. Maximum-likelihood estimation statistics for the ensemble of 484 simulation and recovery experiments reported in Fig. 7. The mean estimate is highlighted by the gray triangle and two observed standard deviations marked gray by lines. The heavy gray ellipse is the 68% confidence region based on the ensemble. The heavy black ellipse is the predicted 68% confidence region based on the covariance predicted from eq. (55), shown and evaluated at the mean estimate, which is a close match for the set.

with the mean marked as a gray triangle, and a cross indicating twice the standard deviation for each of the estimates. In the lower left of each graph we display the percentage of the estimates that falls within the error ellipse calculated for the observed covariance.

Fig. 10 is an alternative graphical rendition of the theoretically predicted and empirically observed parameter-estimate correlation matrix, that is, their covariance matrix normalized to unity along the diagonal. Parameter trade-offs are inherent in the Matérn description (see, for example, eq. 14). The theoretically calculated correlations are $\{\sigma^2, \nu\} = -0.3714$, $\{\sigma^2, \rho\} = 0.8906$, and $\{\nu, \rho\} = -0.6701$, which are good predictors for those observed, as the empirical values are $\{\sigma^2, \nu\} = -0.4195$, $\{\sigma^2, \rho\} = 0.8772$, and $\{\nu, \rho\} = -0.7259$, respectively.

4.6 Analysis of residuals

The terms $\mathcal{S}_{\boldsymbol{\theta}}^{-1}(k)\,|H(\mathbf{k})|^2$ and $\bar{\mathcal{S}}_{\boldsymbol{\theta}}^{-1}(\mathbf{k})\,|H(\mathbf{k})|^2$ that have appeared above in the expressions for the likelihoods (39), (48), and their derivatives (40) and (52) contain the ratio of the observed periodogram of the data to the (blurred) spectral density predicted under the model. Since the spectral density is a 'scale' parameter (as opposed to a 'location') this ratio has the usual interpretation as a measure of misfit. We

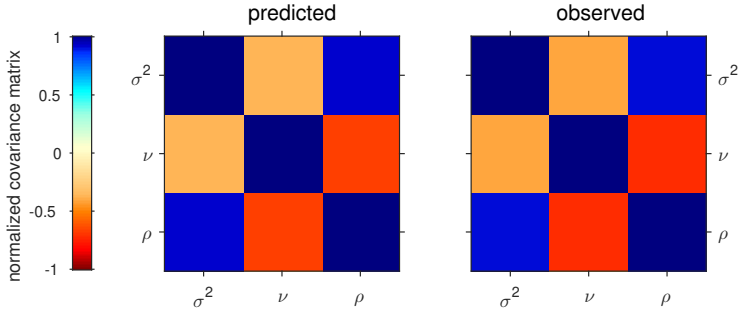

Figure 10. Comparison of the covariance predicted via eq. (55) and the covariance observed on the set of experiments reported in Figs 7 and 9. Shown are the relevant correlation matrices between the estimators for the three Matérn parameters σ^2, ν, and ρ, highlighting the relatively strong trade-off between ν and ρ.

have previously noted that if the Gaussian model fits, the expectation of this quadratic is $\langle \bar{\mathcal{S}}_{\boldsymbol{\theta}}^{-1}(\mathbf{k})\,|H(\mathbf{k})|^2\rangle = 1$. Here we follow Simons & Olhede (2013) to maintain that twice this quantity, a quadratic form of Gaussian random variables, given that $\mathcal{H}(\mathbf{x})$ and thus $H(\mathbf{k})$ are assumed to be Gaussian and uncorrelated, should be a chi-squared random variable with two degrees of freedom,

$$X_{\boldsymbol{\theta}}(\mathbf{k}) = \bar{\mathcal{S}}_{\boldsymbol{\theta}}^{-1}(\mathbf{k})\,|H(\mathbf{k})|^2 \sim \chi^2_2/2. \tag{65}$$

Equipped with this knowledge we can examine how closely the ratios $X_{\boldsymbol{\theta}}(\mathbf{k})$, i.e., the 'residuals', follow the distribution (65), and use the match or lack thereof as a basis to accept or reject the model that the data are indeed given by a Matérn process of the specified parameters.

It is imprudent to ignore and impossible to overstate the importance of such a hypothesis test. Apart from serious numerical instability and potential run-away effects, possibly caused by improper initialization of or unrealistic constraints on the optimization procedure, maximum–likelihood inversion will always return the parameter set *with maximum likelihood*. But whether the most likely model is, in fact, *any* good, then remains to be ascertained. Establishing whether eq. (65) in fact holds can be carried out visually, by inspection of the overlay of the histograms of $2X_{\boldsymbol{\theta}}$ across all wave vectors with the probability density function χ^2_2, and by making 'quantile-quantile' plots of the ranked values of $2X_{\boldsymbol{\theta}}$ versus the inverse cumulative density function of χ^2_2 evaluated at their corresponding fractional ranks. Moreover, the two-dimensional map of $X_{\boldsymbol{\theta}}(\mathbf{k})$ should show no residual structure, and will contain information on possible wavenumber ranges or specific directions (for example, in the presence of anisotropy) in which the data might be over- or under-fit. All three such representations of model quality (histograms, quantile-quantile 'Q-Q' plots, and two-dimensional wavenumber maps of the residuals) must be thoroughly scrutinized.

Beyond visual inspection, it is desirable to design a formal test for when the hypothesis of isotropic Matérn behavior needs to be abandoned, and the veracity of the parameters recovered by likelihood maximization called into question, regardless of how narrow their uncertainty intervals based on eq. (55) may be. Failing the test could be due, for example, to non-Gaussianity, or to the presence of patterns or preferred directions indicating that the data should rather be interpreted under anisotropic (e.g., Goff & Jordan 1989b; Herzfeld & Overbeck 1999; Olhede 2008; Olhede et al. 2014) extensions of the model. We save developing alternative hypotheses for future work.

For a given modeled data sample, we propose as a test statistic the mean-squared deviation from the expected value of the residual ratio,

$$s^2_X = \frac{1}{MN}\sum_{\mathbf{k}}[X_{\boldsymbol{\theta}}(\mathbf{k}) - 1]^2. \tag{66}$$

The smoothness and boundedness of the spectrum $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$, the presumed independence of $X_{\boldsymbol{\theta}}(\mathbf{k})$ between wavenumbers, and the central limit theorem should help the variable s^2_X to converge to a normal variate. The central moments of the pth power of chi-squared variables with m degrees of freedom (Davison 2003) satisfy $\big\langle \big[\chi^2_m\big]^p \big\rangle = 2^p\Gamma(p+m/2)/\Gamma(m/2)$, from which we obtain

$$\big\langle X^p_{\boldsymbol{\theta}} \big\rangle = \Gamma(p+1) = p!. \tag{67}$$

The case $p = 1$ discussed by Simons & Olhede (2013) in reducing eq. (42) to eq. (43), is easily verified. Evaluating eq. (67) for the case $p = 2$ then yields the expectation of eq. (66), our test statistic,

$$\langle s^2_X \rangle = \frac{1}{MN}\sum_{\mathbf{k}} \big\langle X^2_{\boldsymbol{\theta}}(\mathbf{k}) - 2X_{\boldsymbol{\theta}}(\mathbf{k}) + 1 \big\rangle = 1. \tag{68}$$

For its variance, again assuming independence between the wave vectors, we find from elementary calculations that

$$MN\mathrm{var}\big\{s^2_X\big\} = \mathrm{var}\big\{X^2_{\boldsymbol{\theta}}\big\} + 4\,\mathrm{var}\big\{X_{\boldsymbol{\theta}}\big\} - 4\,\mathrm{cov}\big\{X^2_{\boldsymbol{\theta}}, X_{\boldsymbol{\theta}}\big\} \tag{69}$$

$$= \big\langle X^4_{\boldsymbol{\theta}}\big\rangle - \big\langle X^2_{\boldsymbol{\theta}}\big\rangle^2 + 4\big\langle X^2_{\boldsymbol{\theta}}\big\rangle - 4\big\langle X_{\boldsymbol{\theta}}\big\rangle^2 - 4\big\langle X^3_{\boldsymbol{\theta}}\big\rangle + 4\big\langle X^2_{\boldsymbol{\theta}}\big\rangle\big\langle X_{\boldsymbol{\theta}}\big\rangle = 8. \tag{70}$$

Hence we deduce that our chosen metric converges 'in law' (Ferguson 1996) to a variable distributed as:

$$s^2_X \xrightarrow{\text{law}} \mathcal{N}(1, 8/[MN]). \tag{71}$$

In other words, by computing eq. (66) after finding the maximum-likelihood estimates for the Matérn parameters of a data set, we are in a position to test whether the residuals are distributed according to the theory, rejecting the model at whichever confidence level we envisage.

Fig. 11 enlightens us in this regard. The three panels (histograms, Q-Q plots, 2-D map) show the result of a successful experiment with parameters similar to those of the runs presented in Fig. 7, in which the Matérn parameters were very well recovered. The residuals $X_{\boldsymbol{\theta}}(\mathbf{k})$ show the expected distributional behavior without any hint of remaining structure, privileged directions or otherwise. The sample mean and the sample variance of the variable $X_{\boldsymbol{\theta}}$ are listed above the first panel. Per eq. (65), both are expected to be one. Above the second panel are the test statistic s^2_X, its variance under the null hypothesis, the decision to accept (in this case) or reject, and the two-sided probability that values more extreme than the calculated one are likely to occur under the model. All of these entities factor into our decision making.

Fig. 12 validates the distribution of the test statistic through a series of simulation experiments with the same parameters as those used for Fig. 11. Reporting on the 'test of the test', specifically, whether eq. (71) holds, the two panels illustrate the distribution of s^2_X across 175 realizations. Sample mean and variance of the test statistic are labeled above the first panel, which displays the histogram of s^2_X over the simulations. The second panel shows the linearity of the Q-Q plots. Our conclusion is that using s^2_X as a statistic results in a useful and sensitive test on the appropriateness of the Matérn model, whatever its parameters, and irrespectively of their confidence intervals.

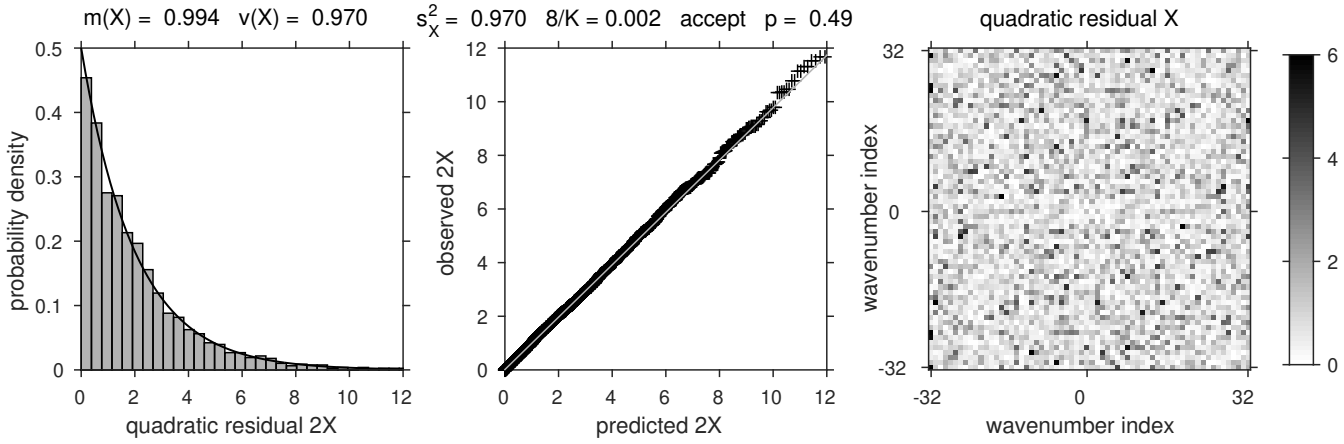

Figure 11. Residual statistics of simulations carried out on a 101×111 grid, with spacings $\Delta x = \Delta y = 10$ km, convolutional blurring, and true Matérn parameter values of $\sigma_0^2 = 1$ km^2, $\nu_0 = 2.5$, $\rho_0 = 20$ km. Results pertaining to one of the simulations and its maximum-likelihood recovery. Distribution of the variable $X_{\boldsymbol{\theta}}(\mathbf{k})$ of eq. (65), as a histogram across all wavenumbers with the theoretical distribution superposed, as a quantile-quantile plot for the distribution in question, and as a spectral-domain map.

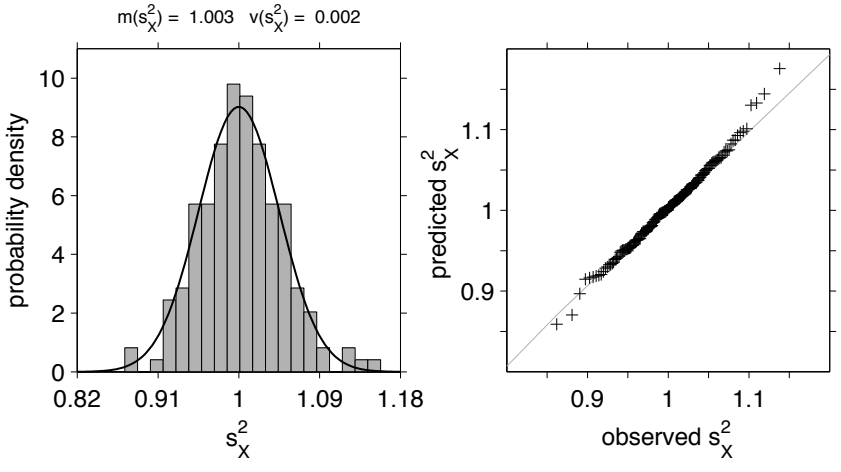

Figure 12. The behavior of the test statistic for appropriateness of the Matérn model, s^2_X of eq. (66), across an ensemble of 175 simulation and recovery experiments. Histogram and its prediction, and quantile-quantile plot comparing observations to predictions. The theoretical behavior of eq. (71) is validated.

5 MAXIMUM-LIKELIHOOD PRACTICE

Our fundamental point of departure is that every geological or geophysical spatial data set is a sample, a realization, of a parameterizable random process, whose parameters we seek to recover. Whether these statistical parameters then are directly interpretable (e.g., Malinverno 1991; Whittaker et al. 2008), used for interpolation or extrapolation (e.g., Mareschal 1989; Sandwell et al. 2022), anomaly detection and classification (e.g., Herzfeld & Higginson 1996; Herzfeld & Zahner 2001), or in order to infer from them any kind of other, e.g., geophysical, oceanographic, or geological, information (e.g., Stephenson 1984; Goff & Arbic 2010; Song & Simons 2003; Persson 2006; Sagy et al. 2007; Bottero et al. 2020; Cristini & Komatitsch 2012; Grainger et al. 2021; Wunsch 2022) is beyond the immediate remit of this paper.

There may not appear to be a dearth of approaches to solve problems of the nature described above, but in remote-sensing, geology, geophysics, oceanography, and planetary science, many methods in common use yield results that are biased, too variable, or too vague, to be of quantitative statistical use. An example to help clarify the needs addressed in this paper is how best to answer the seemingly straightforward question (e.g., Munk 1955; Kreslavsky & Head 2000; Shepard et al. 2001; Grohmann et al. 2011; Candela et al. 2012; Reich et al. 2013; Guérin-Marthe et al. 2023): "how *rough* is that (topographic, ocean-floor, sea-surface, earthquake-fault, rock-sample, ... tooth) surface?"

The overarching accomplishment of this paper is that we have developed a fast, efficient, and effective strategy for such parameter estimation problems, for Gaussian processes governed by a flexible class of two-point covariance structures, under realistic sampling scenarios which require explicit consideration and treatment in order to render the solutions interpretable, intercomparable, and robust, with quantifiable uncertainties and a test for model appropriateness. The sampling patterns discussed so far, and in the remainder of this paper, are regularly spaced Cartesian rectangular grids. Should the boundaries of a region be irregular, or the sampling be incomplete, as long as the selection function or window $w(\mathbf{x})$ in eq. (25) can accommodate picking out what is actually being observed, our theory and methods will continue to hold. We will return to incomplete, be it random, deterministic, or otherwise structured sampling patterns in future work.

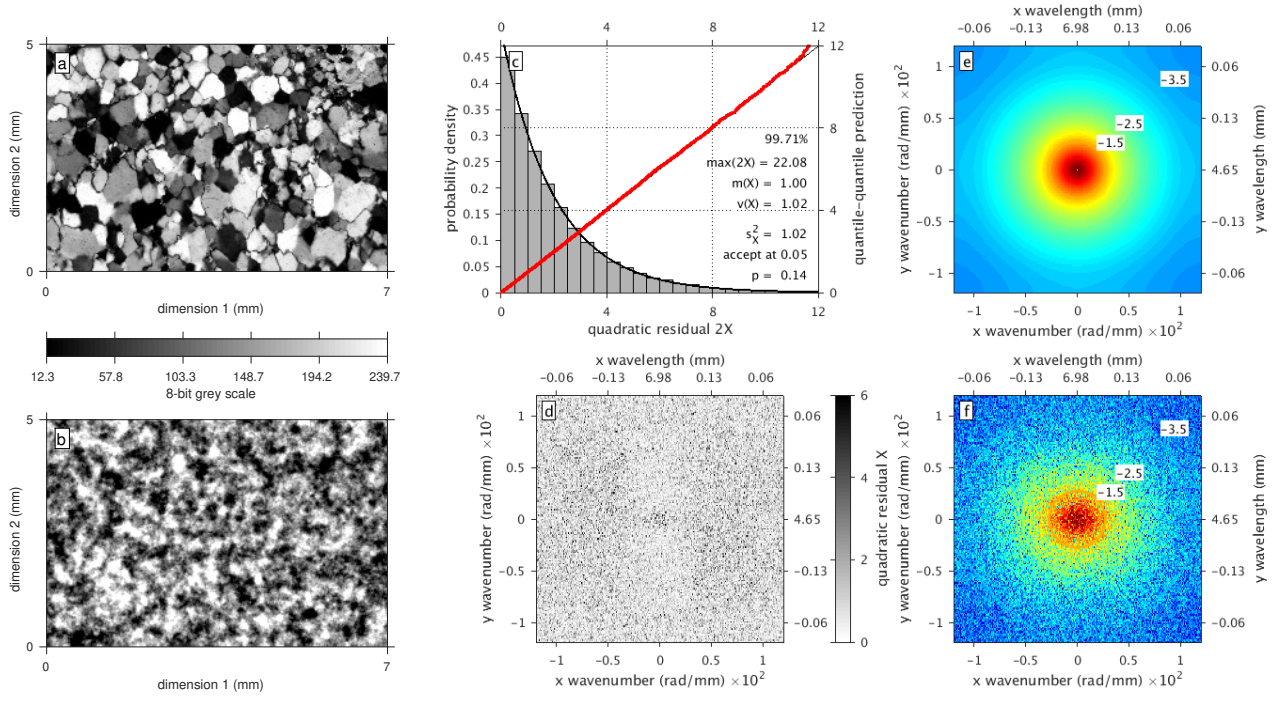

Figure 13. Maximum-likelihood analysis for the Matérn covariance structure of a thin section of a quartzite rock under cross-polarized light (Da Mommio & Pease 2025). Shown are the observed data (*a*) and a synthetic randomly generated from the Matérn parameters recovered as our estimate (*b*) as discussed in Sec. 4.3, a histogram and a quantile-quantile (Q-Q) plot of the quadratic residual $2X_{\boldsymbol{\theta}}$ (*c*), with $X_{\boldsymbol{\theta}}(\mathbf{k})$ being rendered and inspected for patterns in wave vector space (*d*), as discussed in Sec. 4.6. Also shown are the expected (*e*) and observed (*f*) periodograms, $\tilde{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$ and $|H(\mathbf{k})|^2$ respectively, with contour lines (labeled by their exponent) for the former overlain on the latter. The model fits extremely well over the entire wavenumber range.

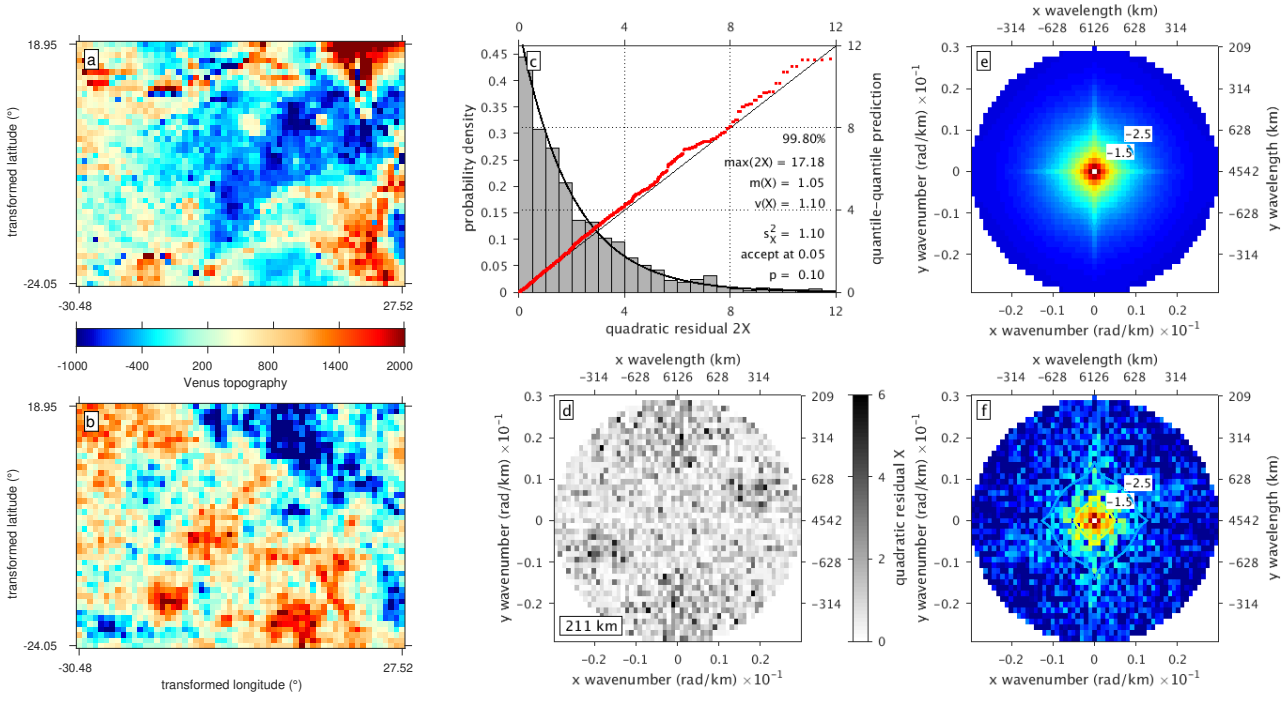

Figure 14. Maximum-likelihood analysis for the Matérn covariance structure of a patch of Venus topography. Layout and annotation are as in Fig. 13. To counter the isotropic filtering present in the original data set (Rappaport et al. 1999; Eggers 2013), we calculated the likelihood and based our decision solely on the wavenumbers within the rendered spectral disk. The model fits very well, small residual departures visible in the Q-Q plot notwithstanding, which are indeed expected for correlated order-statistics at both ends (Davison 2003). We draw attention to the numerical annotations in the top middle panel. The percentile corresponding to the horizontal axis limit is shown. Below that, maximum, mean, and variance of the residual $X_{\boldsymbol{\theta}}$ (see eq. 65), the test statistic s^2_X (see eq. 66), the decision at the 95% significance level, and the one-sided exceedance probability for the test of it being normally distributed (see eq. 71).

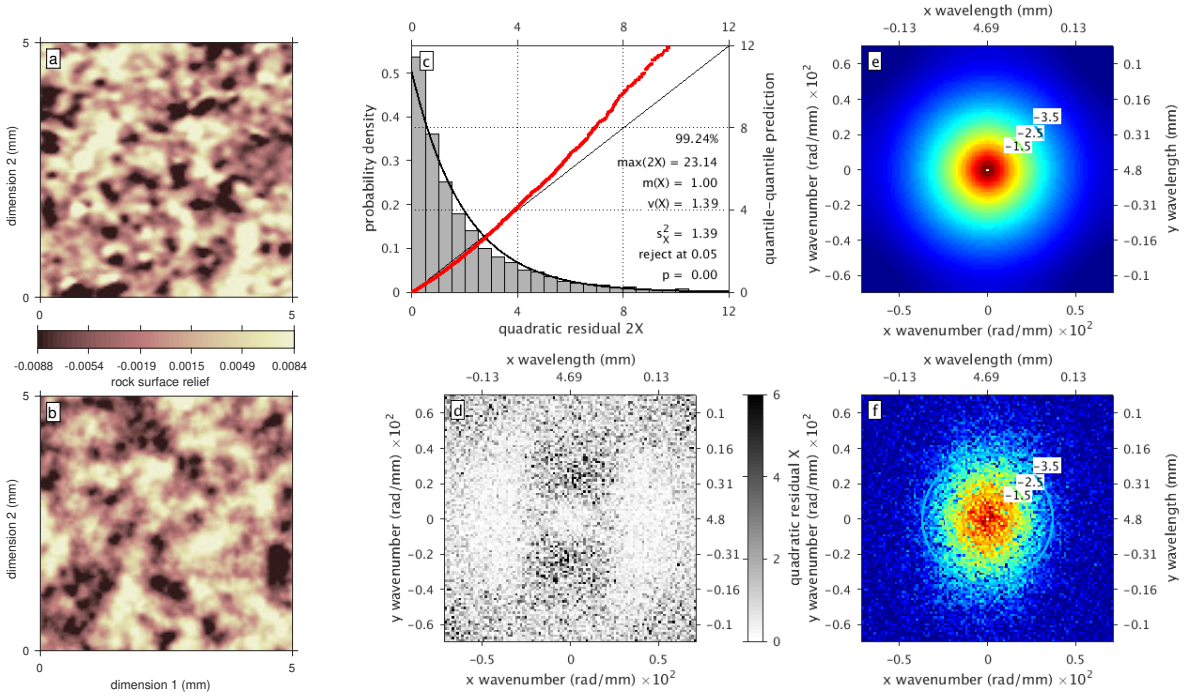

Figure 15. Maximum-likelihood analysis for the Matérn covariance structure of a saw-cut, polished, and sandblasted granitic rock surface fabricated in a friction experiment (Guérin-Marthe et al. 2023). Layout and annotation are as in Figs 13–14. The model fits relatively poorly under the most intransigent interpretation of the Q-Q plot, which could be argued is too demanding, as the 'mass' of the distribution is reasonably well aligned with statistical expectation. The residuals show relatively strong hints of unmodeled structure with directional and wavenumber dependence. On the whole, we consider this to be an acceptable model, and in future work we might relax the strict adherence to eq. (71) as a make-or-break decision for model acceptance.

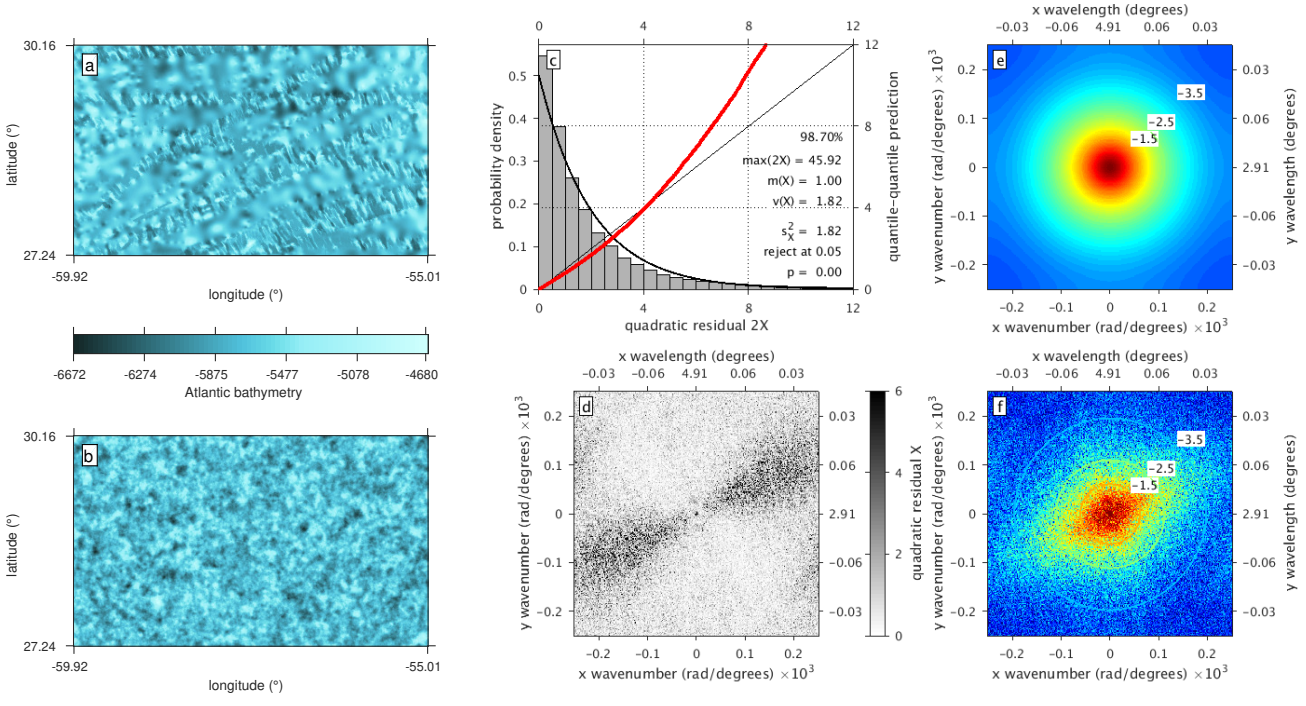

Figure 16. Maximum-likelihood analysis for the Matérn covariance structure of Atlantic seafloor bathymetry. Layout and annotation as in Figs 13–15. Of all the examples shown so far, this model fits most poorly, while it clearly still has much merit. However, its failure to be accepted based on the stringent eq. (71) will drive us to consider models beyond the isotropic stationary Matérn class. The strong directional dependence of the residuals reveals that anisotropic behavior will need to be considered, as is indeed expected both based on the intrinsic geophysical and geological properties of the seafloor (Goff & Jordan 1989b), and from the heterogeneous and anisotropic model (GEBCO Bathymetric Compilation Group 2019) built from compiled multiresolution data sources that include both direct shipboard seafloor sounding and interpreted satellite altimetry sea-surface measurements (Smith & Sandwell 1997).

The statistical literature abounds with methods and examples of asymptotic (infill, growing, and mixed-domain) convergence results for specific special cases of the Matérn spectral density or covariance, that is, for specific values of ν in eqs (12) and (15), which resolve to simpler forms (Guttorp & Gneiting 2006). Sec. 4.3 showed how the statistical behavior of the solutions depends both on the sampling and on the random process being sampled, and it therefore must remain beyond the scope of this paper to discuss this interplay with any aspiration to generality. The important cases of anisotropy, non-stationarity, non-Gaussianity, and correlations between multivariable fields will await further treatment under our framework. Any and all of those considerations, however, will rely on a thorough understanding of the null-hypothesis of univariate, stationary, isotropic, Gaussian fields such as are provided here.

For a first confrontation of our new results with geophysical and geological practice, we will limit ourselves to showing a few illuminating examples of parameter recovery, uncertainty quantification, and model testing, on completely sampled, rectangular grids. In Figs 13–16, the left column contains an image of the data, underneath which we plot a simulation based on the Matérn model with the parameters derived from our procedure. The middle column appears in a layout that is a combination of the three panels of Fig. 11, showing the histogram of $X_{\boldsymbol{\theta}}(\mathbf{k}) = \bar{\mathcal{S}}_{\boldsymbol{\theta}}^{-1}(\mathbf{k})\,|H(\mathbf{k})|^2$ of eq. (65), its theoretical probability density (using the left ordinate) and Q-Q plot (using the right ordinate), the test statistic s_X^2 of eq. (66) and the decision based on it via eq. (71). In the panel below we show the values of $X_{\boldsymbol{\theta}}(\mathbf{k})$ in wave vector space, to inspect for any undesirable patterns that may persist. The panels in the rightmost column are as the rightmost top panels in Fig. 2, showing the blurred spectral density $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(\mathbf{k})$, that is, the expectation of the windowed periodogram, $\langle|H(\mathbf{k})|^2\rangle$, and the periodogram $|H(\mathbf{k})|^2$ observed from the data. Contours of the expected spectrum are drawn on the observed, highlighting where they match.

The optimization was carried out by a simple unconstrained finite-difference gradient-based method (using MATLAB's `fminunc`). Starting with the sample variance as a candidate for σ^2, our runs were initialized with values that, while inspired by comments made by Vanmarcke (1983) relative to ρ, and Sykulski et al. (2019) relative to ν, have not needed to be more than "reasonable" guesses. Up to a planar trend was removed for the data prior to analysis, and a cosine-squared window tapering 10% all around was applied to smooth any edges, exactly as we had done to make Figs 11–12. See Table 1 for a summary of the results discussed next.

The first example is a photomicrograph of a rock sample in a 30 μm thin-section used for optical petrography. Fig. 13 is a gray-scale image of a quartzite, a metamorphic rock composed almost entirely of the mineral quartz (SiO_2). The field of view is 7 mm, and the image was taken under cross-polarized light (Da Mommio & Pease 2025), revealing individual grains with characteristic undulose extinction. The Matérn model fits very well in this case, despite the sand grains being distinct entities with relatively constant grayscale values and visible grain boundaries. The synthetic will not be mistaken for an actual, possibly non-Gaussian, rock sample, yet it remains statistically indistinguishable from the perspective of Gaussian field modeling. Virtually no discernible patterns characterize the spectral-domain residuals, and the power-spectral densities are almost entirely isotropic, as expected for this type of uniformly recrystallized metamorphic rock.

The second example is a patch of topography from the planet Venus. Fig. 14 shows the region around Hinemoa Planitia obtained from radar altimetry mapping by the Magellan spacecraft (Rappaport et al. 1999), expanded from spherical harmonics, reprojected to center on $(261.57^\circ, 5.29^\circ)$, and downsampled from the original full resolution (Eggers 2013). The area shown occupies a mean elevation of 150 m above the mean planetary radius, with a standard deviation of 756 m. The rendered colors span the 1st through the 99th percentile, which cover relative elevations 1161 m below and 2176 m above the deplaned surface. To accommodate an isotropic Nyquist filtering step applied to the original data, which effectively erased all power from the corners outside the circle inscribed in the Cartesian wave vector rectangle, we restricted the likelihood calculation to the disk-shaped domain of radius 211 km shown in the spectral plots. The Matérn model fits the data well and is formally accepted as a null-hypothesis. Nevertheless, there are visible hints of systematic departures from complete isotropy in the middle- to large-wavenumber range that may well prove to be geologically significant upon further analysis. Furthermore, by our definition (Eggers 2013), Hinemoa Planitia as a coherent geological unit is an irregularly bounded subdomain of the rectangular patch shown, which will require applying spatially selective windowing to fully and uniquely characterize as a locally stationary random field.

The third example is a portion of a granite surface taken before a tribological experiment (Guérin-Marthe et al. 2023). Fig. 15 shows a piece of the relief on La Peyratte granite "sample R1 top", measured using a Keyence VR-3200 surface scanning light microscope, with a reported root-mean-squared "asperity" height of 12–16 μm. The "rough" surface, created by sandblasting the saw-cut, pre-polished, sample surface with silicon carbide (SiC) powder is marked by long-wavelength relief and high gradients between extreme peaks and troughs which proved challenging to capture in the modeled field. Hence we deplaned the image and winsorized it within the 7.5 and 92.5 percentiles prior to analysis. The results, while we deem the synthetic visually encouraging, nevertheless warrant rejection of the Matérn model. The spectral-domain residuals fail to meet our strict distributional criterion, and show systematic deviation from what should be a complete lack of structure in wave vector space. Despite these misgivings, the best-fit solutions for the Matérn parameters made on successive measurements of rock surface relief as it gets ground down and smoothed in fault-friction experiments hold promise for their objective if imperfect characterization.

The fourth and last example involves a portion of the seafloor in the North Atlantic. Fig. 16 shows the GEBCO Bathymetric Compilation Group (2019) bathymetric model for the area, which lies at a mean elevation of 5738 m below the sea surface, with a standard deviation of 271 m. The rendered colors span the depths between 6672 m and 4680 m. Oceanic bathymetry is a *model* based on few direct observations: swaths of high-resolution shipboard-derived multibeam sonar data draped over a long-wavelength low-resolution model derived from radar altimetry of the sea surface, which gravitationally reflects the topography of the ocean bottom (Smith & Sandwell 1997). Whether completely sampled or not, ocean bathymetry, as a mirror of plate tectonics, rarely if ever shows isotropic behavior. The isotropic Matérn model developed in this paper clearly does not apply. We include this example because our analysis allows for the testing (and rejection, as in this case) of isotropic behavior, and the initial evaluation of anisotropy, for which suitable models and estimation methods need to be developed.

Example	Figure	K	σ^2/s^2	$\pm\%$	ν	$\pm\%$	ρ	$\pi\rho/r$	$\pm\%$	$\{\sigma^2,\nu\}$	$\{\sigma^2,\rho\}$	$\{\nu,\rho\}$
Seafloor	16	92196	0.92	3	1.26	0.6	1.75 km	0.01	2	-38	$+90$	-72
Quartzite	13	47526	0.91	4	0.91	1.0	0.0512 mm	0.02	3	-31	$+90$	-67
Granite	15	11556	1.64	10	1.64	1.4	0.1033 mm	0.05	4	-39	$+92$	-71
Venus	14	2026	1.64	53	0.31	9.0	841 km	0.34	96	-22	$+96$	-46

Table 1. Results from our experiments with geological and geophysical data, sorted by the degrees of freedom of the analysis, the size of the field, $K = M \times N$, adjusted to the wavenumber disk for the case of Venus. The estimate for the variance σ^2 is quoted as a fraction of the sample variance of the data set, here denoted s^2. In the presence of significant range, s^2 is small relative to σ^2. The estimate of the range parameter ρ is first presented in units of the field and then is multiplied by π, see Fig. 1, and expressed as a fraction of r, the length of the diagonal of the data grid. For all three parameters, the one-standard-deviation estimation uncertainty is listed to the nearest per cent of the parameter itself. The final three columns contain the correlation between parameter estimates, as in Fig. 10, in per cent.

6 CONCLUSIONS

The Matérn (1960) covariance provides a parsimonious, three-parameter, characterization of stationary and isotropic geological and geophysical univariate Gaussian fields. While it has enjoyed choice application over the years, no truly pragmatic framework existed for the analysis of finite, sampled spatial data sets, including those with irregular boundaries or missing grid points. To recover the diagnostic Matérn parameters of variance, smoothness, and range, exact space-based likelihood methods are almost always impractical or computationally out of reach. Spectral-domain likelihood approximations introduced by Whittle (1954) hold the advantage of computational and statistical efficiency, but their parameter estimation bias has been an ongoing cause for concern, especially for incompletely and irregularly bounded sampled fields. The debiased spatial Whittle likelihood of Guillaumin et al. (2022) correctly accounts for edge effects and data omission by blurring the spectral density function in a manner that is both exact and fast. We showed its efficacy in recovering Matérn covariance parameters on synthetically generated and actually observed geophysical fields. A quasi- or pseudo-likelihood, the debiased Whittle likelihood is indifferent to wave vector correlations during parameter estimation. We showed how to calculate these interactions, exactly, and how to include them to derive the full parameter estimation covariance matrix. Only when these are properly acknowledged do we produce reliable uncertainty estimates, as we validated experimentally. Our final addition to complete the theory involves a test for the epistemic appropriateness of the Matérn model, against which possible non-stationary, an-isotropic, non-Gaussian departures from the null-hypothesis can be evaluated.

Our work allows us to ask and confidently answer questions related to the statistical description of geological and geophysical fields and processes in a manner that produces solutions that are interpretable, intercomparable, and robust, with quantifiable uncertainties and tests for overall model fitness. What numbers capture the essence of a patch of spatial data? Can we estimate them, can we derive uncertainty bounds for them, can we simulate "new" realizations of fields that behave exactly as if they were derived from similar processes? The answer to all of these questions is "yes". Our approach holds in one, two, and three dimensions, and generalizes to multiple variables, e.g., when topography and gravity are being considered jointly (e.g., linked by flexure, erosion, or other surface and sub-surface modifying processes, see Simons & Olhede 2013), or for remote-sensing applications. While currently agnostic as to the application domain, our ultimate goal is to derive geophysical "process" from statistical "parameters", e.g., to be able to assign likely formation mechanisms and histories for the data under consideration, with specific input from domain knowledge.

7 DATA AVAILABILITY

All the code used to conduct the calculations and produce the figures in this paper made in MATLAB by the authors is openly available as Release 2.0.0 from `https://github.com/csdms-contrib/slepian_juliet`, doi: `10.5281/zenodo.4085253`. For an alternative code base in Python, see also Guillaumin et al. (2026). Also see the *R* package by Paciorek (2007). The quartzite photomicrograph is described by (Da Mommio & Pease 2025) and was obtained from `https://www.alexstrekeisen.it/english/meta/quartzite.php`, and used with permission from Dr. Alessandro Da Mommio (`alexdm83@libero.it`). The original Venus topography model `shtjv360` was constructed by Rappaport et al. (1999) and resides at `https://pds-geosciences.wustl.edu/missions/magellan/shadr_topo_grav/index.htm`. The granite surface data pertain to the paper by Guérin-Marthe et al. (2023) and are available from `https://zenodo.org/records/6411819`. The seafloor bathymetry model `GEBCO_2019` is available from `https://www.gebco.net/data-products/gridded-bathymetry-data/gebco-2019` as described by GEBCO Bathymetric Compilation Group (2019).

8 ACKNOWLEDGEMENTS

This work was sponsored by the National Aeronautics and Space Administration under grant NNX11AQ45G. OLW gratefully acknowledges financial support from the Schmidt DataX Fund, grant number 22-008, at Princeton University, made possible through a major gift from the Schmidt Futures Foundation. FJS thanks the Institute for Advanced Study for a stimulating research environment during 2023–2024, and KU Leuven and the Groot Begijnhof for enabling productive Summer months. We thank Erin O'Neil for helpful comments on the final draft. We are grateful to the Associate Editor Carl Tape, and to John Goff and Michael Stein for constructive reviews that improved the manuscript.

REFERENCES

Abramowitz, M. & Stegun, I. A., 1965. *Handbook of Mathematical Functions*, Dover, New York.

Adler, R. J., 1981. *The Geometry of Random Fields*, Wiley, New York.

Aharonson, O., Zuber, M. T., Neumann, G. A. & Head, J. W., 1998. Mars: Northern hemisphere slopes and slope distributions, *Geophys. Res. Lett.*, **25**(24), 4413–4416.

Appourchaux, T., Gizon, L. & Rabello-Soares, M.-C., 1998. The art of fitting p-mode spectra. I. Maximum likelihood estimation, *Astron. Astroph. Suppl. Ser.*, **132**, 107–119, doi: 10.1051/aas:1998441.

Appourchaux, T., Rabello-Soares, M.-C. & Gizon, L., 1998. The art of fitting p-mode spectra. II. Leakage and noise covariance matrices, *Astron. Astroph. Suppl. Ser.*, **132**, 121–132, doi: 10.1051/aas:1998440.

Aster, R. C., Borchers, B. & Thurber, C. H., 2018. *Parameter Estimation and Inverse Problems*, Elsevier Academic Press, San Diego, Calif., 3rd edn.

Baig, A. M. & Dahlen, F. A., 2004. Statistics of traveltimes and amplitudes in random media, *Geophys. J. Int.*, **158**(1), 187–210, doi: 10.1111/j.1365–246X.2004.02300.x.

Baig, A. M., Dahlen, F. A. & Hung, S.-H., 2003. Traveltimes of waves in three-dimensional random media, *Geophys. J. Int.*, **153**(2), 467–482, doi: 10.1046/j.1365–246X.2003.01905.x.

Becker, T. W., Browaeys, J. T. & Jordan, T. H., 2007. Stochastic analysis of shear-wave splitting length scales, *Earth Planet. Sci. Lett.*, **259**(3–4), 526–540, doi: 10.1016/j.epsl.2007.05.010.

Bendat, J. S. & Piersol, A. G., 2000. *Random Data: Analysis and Measurement Procedures*, John Wiley, New York, 3rd edn.

Botev, Z. I., Grotowski, J. F. & Kroese, D. P., 2010. Kernel density estimation via diffusion, *Ann. Stat.*, **38**(5), 2916–2957, doi: 10.1214/10–AOS799.

Bottero, A., Cristini, P. & Komatitsch, D., 2020. On the influence of slopes, source, seabed and water column properties on T waves: Generation at shore, *Pure Appl. Geophys.*, **177**, 5695–5711, doi: 10.1007/s00024–020–02611–z.

Candela, T., Renard, F., Klinger, Y., Mair, K., Schmittbuhl, J. & Brodsky, E. E., 2012. Roughness of fault surfaces over nine decades of length scales, *J. Geophys. Res.*, **117**, B08409, doi: 10.1029/2011JB009041, 2012.

Carpentier, S. & Roy-Chowdhury, K., 2007. Underestimation of scale lengths in stochastic fields and their seismic response: a quantification exercise, *Geophys. J. Int.*, **169**(2), 547–562.

Christakos, G., 1992. *Random Field Models in Earth Sciences*, Academic Press, San Diego, Calif., 2nd edn.

Chu, T., 2023. Mixed domain asymptotics for geostatistical processes, *Statistica Sinica*, **33**(1), 551–571, doi: 10.5705/ss.202020.0092.

Cox, D. R. & Hinkley, D. V., 1974. *Theoretical Statistics*, Chapman and Hall, London, UK.

Cramér, H., 1942. On harmonic analysis in certain fuctional spaces, *Arkiv Mat. Astr. Fysik*, **28B**, 1–7.

Cressie, N., 1993. *Statistics for Spatial Data*, John Wiley, London, UK.

Cristini, P. & Komatitsch, D., 2012. Some illustrative examples of the use of a spectral-element method in ocean acoustics, *J. Acoust. Soc. Am.*, **131**(3), EL229–EL235, doi: 10.1121/1.3682459.

Da Mommio, A. & Pease, V., 2025. *Atlas of Minerals and Igneous and Metamorphic Rocks in Thin-Section*, Cambridge Univ. Press, Cambridge, UK.

Dahlhaus, R., 1984, Parameter estimation of stationary processes with spectra containing strong peaks, in *Robust and Nonlinear Time Series Analysis: Proceedings of a Workshop Organized by the Sonderforschungsbereich 123 "Stochastische Mathematische Modelle", Heidelberg 1983*, pp. 50–67, doi: 10.1007/978-1-4615-7821-5_4, Springer.

Dahlhaus, R. & Künsch, H., 1987. Edge effects and efficient parameter estimation for stationary random fields, *Biometrika*, **74**(4), 877–882, doi: 10.1093/biomet/74.4.877.

Davison, A. C., 2003. *Statistical Models*, Cambridge Univ. Press, Cambridge, UK.

Deb, S., Pourahmadi, M. & Wu, W. B., 2017. An asymptotic theory for spectral analysis of random fields, *Electron. J. Stat.*, **11**, 4297–4322, doi: 10.1214/17–EJS1326.

Dzhaparidze, K. O. & Yaglom, A. M., 1983, Spectrum parameter estimation in time series analysis, in *Developments in Statistics*, edited by P. Krishnaiah, vol. 4, pp. 1–96, doi: 10.1016/B978–0–12–426604–9.50007–0, Academic Press, New York.

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

Ferguson, T. S., 1996. *A Course in Large Sample Theory*, Chapman and Hall/CRC Press, New York.

Fournier, D., Gizon, L., Hohage, T. & Birch, A. C., 2014. Generalization of the noise model for time-distance helioseismology, *Astron. Astroph.*, **567**, A137, doi: 10.1051/0004–6361/201423580.

Fuentes, M., 2007. Approximate likelihood for large irregularly spaced spatial data, *J. Acoust. Soc. Am.*, **102**(477), 321–331, doi: 10.1198/016214506000000852.

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

Geoga, C. J., Marin, O., Schanen, M. & Stein, M. L., 2023. Fitting Matérn smoothness parameters using automatic differentiation, *Stat. Comput.*, **33**(2), 48, doi: 10.1007/s11222–022–10127–w.

Gizon, L. & Birch, A. C., 2004. Time-distance helioseismology: noise estimation, *Astroph. J.*, **614**(1), 472–489, doi: 10.1086/423367.

Gneiting, T., Ševčíková, H. & Percival, D. B., 2012. Estimators of fractal dimension: Asessing the roughness of time series and spatial data, *Stat. Sci.*, **27**(2), 247–277, doi: 10.1214/11–STS370.

Goff, J. A. & Arbic, B. K., 2010. Global prediction of abyssal hill roughness statistics for use in ocean models from digital maps of paleo-spreading rate, paleo-ridge orientation, and sediment thickness, *Ocean Modelling*, **32**(1–2), 36–43, doi: 10.1016/j.ocemod.2009.10.001.

Goff, J. A. & Jordan, T. H., 1988. Stochastic modeling of seafloor morphology: Inversion of Sea Beam data for second-order statistics, *J. Geophys. Res.*, **93**(B11), 13589–13608, doi: 10.1029/JB093iB11p13589.

Goff, J. A. & Jordan, T. H., 1989. Stochastic modeling of seafloor morphology: Resolution of topographic parameters by Sea Beam data, *IEEE J. Ocean. Eng.*, **14**(4), 326–337, doi: 10.1109/48.35983.

Goff, J. A. & Jordan, T. H., 1989. Stochastic modeling of seafloor morphology: A parameterized Gaussian model, *Geophys. Res. Lett.*, **16**(1), 45–48, doi:

10.1029/GL016i001p00045.

Gradshteyn, I. S. & Ryzhik, I. M., 2000. *Tables of Integrals, Series, and Products*, Academic Press, San Diego, Calif., 6th edn.

Grainger, J. P., Sykulski, A. M., Jonathan, P. & Ewans, K., 2021. Estimating the parameters of ocean wave spectra, *Ocean. Eng.*, **229**, 108934, doi: 10.1016/j.oceaneng.2021.108934.

Grainger, J. P., Rajala, T. A., Murrell, D. J. & C., S. O., 2025. Spectral estimation for spatial point processes and random fields, *Biometrika*, pp. asaf089, 10.1093/biomet/asaf089.

Grohmann, C. H., Smith, M. J. & Riccomini, C., 2011. Multiscale analysis of topographic surface roughness in the Midland Valley, Scotland, *IEEE T. Geosci. Remote*, **49**(4), 1200–1213, doi: 10.1109/TGRS.2010.2053546.

Guérin-Marthe, S., Kwiatek, G., Wang, L., Bonnelye, A., Martínez-Garzón, P. & Dresen, G., 2023. Preparatory slip in laboratory faults: Effects of roughness and load point velocity, *J. Geophys. Res.*, **128**(4), e2022JB025511, doi: 10.1029/2022JB025511.

Guðmundsson, Ó., Davies, J. H. & Clayton, R. W., 1990. Stochastic analysis of global traveltime data: mantle heterogeneity and random errors in the ISC data, *Geophys. J. Int.*, **102**(1), 25–43, doi: 10.1111/j.1365–246X.1990.tb00528.x.

Guillaumin, A. P., Sykulski, A. M., Olhede, S. C., Early, J. J. & Lilly, J. M., 2017. Analysis of non-stationary modulated time series with applications to oceanographic surface flow measurements, *J. Time Ser. Anal.*, **38**(5), 668–710, doi: 10.1111/jtsa.12244.

Guillaumin, A. P., Sykulski, A. M., Olhede, S. C. & Simons, F. J., 2022. The debiased spatial Whittle likelihood, *J. R. Stat. Soc., Ser. B*, **84**(4), 1526–1557, doi: 10.1111/rssb.12539.

Guillaumin, A. P., Goodwin, T., Olivia L. Walbert, A. M. S., C.Olhede, S. & Simons, F. J., 2026. DSWL package: a Python implementation of the Debiased Spatial Whittle Likelihood, *J. Open Source Softw.*, **X**(Y), in revision.

Guinness, J., 2019. Spectral density estimation for random fields via periodic embeddings, *Biometrika*, **106**(2), 267–286, doi: 10.1093/biomet/asz004.

Guinness, J. & Fuentes, M., 2017. Circulant embedding of approximate covariances for inference from Gaussian data on large lattices, *J. Comput. Graph. Stat.*, **26**(1), 88–97, doi: 10.1080/10618600.2016.1164534.

Guttorp, P. & Gneiting, T., 2006. Studies in the history of probability and statistics XLIX. On the Matérn correlation family, *Biometrika*, **93**(4), 989–995, doi: 10.1093/biomet/93.4.989.

Guyon, X., 1982. Parameter estimation for a stationary process on a d-dimensional lattice, *Biometrika*, **69**(1), 95–105, doi: 10.1093/biomet/69.1.95.

Hamilton, A. J. S., 2009, Power spectrum estimation. I. Basics, in *Data Analysis in Cosmology*, edited by V. Martinez, E. Saar, E. Gonzales, & M. Pons-Borderia, vol. 665 of **Lecture Notes in Physics**, pp. 415–431, doi: 10.1007/978–3–540–44767–2_12, Springer, Berlin, Germany.

Hamilton, A. J. S., 2009, Power spectrum estimation II. Linear maximum likelihood, in *Data Analysis in Cosmology*, edited by V. Martinez, E. Saar, E. Gonzales, & M. Pons-Borderia, vol. 665 of **Lecture Notes in Physics**, pp. 433–456, doi: 10.1007/978–3–540–44767–2_13, Springer, Berlin, Germany.

Handcock, M. S. & Stein, M. L., 1993. A Bayesian analysis of kriging, *Technom.*, **35**(4), 403–410, doi: 10.1080/00401706.1993.10485354.

Handcock, M. S. & Wallis, J. R., 1994. An approach to statistical spatial-temporal modeling of meteorological fields, *J. Acoust. Soc. Am.*, **89**(426), 368–378.

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**(1), 97–108, doi: 10.1093/biomet/57.1.97.

Herzfeld, U. & Zahner, O., 2001. A connectionist-geostatistical approach to automated image classification, applied to the analysis of crevasse patterns in surging ice, *Comput. Geosci.*, **27**, 499–512, doi: 10.1016/S0098–3004(00)00089–3.

Herzfeld, U. C. & Higginson, C. A., 1996. Automated geostatistical seafloor classification —– Principles, parameters, feature vectors, and discrimination criteria, *Comput. Geosci.*, **22**(1), 35—-52, doi: 10.1016/0098–3004(96)89522–7.

Herzfeld, U. C. & Overbeck, C., 1999. Analysis and simulation of scale-dependent fractal surfaces with application to seafloor morphology, *Comput. Geosci.*, **25**(1), 979–1007, doi: 10.1016/S0098–3004(99)00062–X.

Herzfeld, U. C., Kim, I. I. & Orcutt, J. A., 1995. Is the ocean floor a fractal?, *Math. Geol.*, **27**(3), 421–462, doi: 10.1007/BF02084611.

Hofmann-Wellenhof, B. & Moritz, H., 2006. *Physical Geodesy*, Springer, New York, 2nd edn.

Hosoya, Y. & Taniguchi, M., 1982. A central limit theorem for stationary processes and the parameter estimation of linear processes, *Ann. Stat.*, pp. 132–153, doi: 10.1007/978-1-4615-7821-5_4.

Isserlis, L., 1918. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables, *Biometrika*, **12**(1–2), 134–139, doi: 10.2307/2331932.

Journel, A. G. & Huijbregts, C. J., 1978. *Mining Geostatistics*, Academic Press, San Diego, Calif.

Katzfuss, M. & Guinness, J., 2021. A general framework for Vecchia approximations of Gaussian processes, *Stat. Sci.*, **36**(1), 124–141, doi: 10.1214/19–STS755.

Kent, J. T. & Mardia, K. V., 1996. Spectral and circulant approximations to the likelihood for stationary Gaussian random fields, *J. Stat. Plann. Infer.*, **50**, 379–394, doi: 10.1016/0378–3758(95)00065–8.

Kitanidis, P. K. & Lane, R. W., 1985. Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method, *J. Hydrol.*, **79**(1), 53–71, doi: 10.1016/0022–1694(85)90181–7.

Kreslavsky, M. A. & Head, J. W., 2000. Kilometer-scale roughness of Mars: Results from MOLA data analysis, *J. Geophys. Res.*, **105**(E11), 26695–26711, doi: 10.1029/2000JE001259.

Kroese, D. P. & Botev, Z. I., 2015, Spatial process simulation, in *Stochastic geometry, spatial statistics and random fields*, edited by V. Schmidt, chap. 12, pp. 369–404, doi: 10.1007/978–3–319–10064–7_12, Springer, Heidelberg, Germany.

Lambeck, K., 1988. *Geophysical Geodesy*, Oxford Univ. Press, New York.

Landais, F., Schmidt, F. & Lovejoy, S., 2019. Multifractal topography of several planetary bodies in the solar system, *Icarus*, **319**, 14–20, doi: 10.1016/j.icarus.2018.07.005.

Lilly, J. M., Sykulski, A. M., Early, J. J. & Olhede, S. C., 2017. Fractional Brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion, *Nonlin. Proc. Geophys.*, **24**(3), 481–514, doi: 10.5194/npg–24–481–2017.

Lindgren, F., Rue, H. & Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, *J. R. Stat. Soc., Ser. B*, **73**(4), 423–498, doi: 10.1111/j.1467–9868.2011.00777.x.

Malinverno, A., 1991. Inverse square-root dependence of mid-ocean-ridge flank roughness on spreading rate, *Nature*, **352**, 58–60, doi: 10.1038/352058a0.

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

doi: 10.2307/2336405.

Mareschal, J.-C., 1989. Fractal reconstruction of sea-floor topography, *Pure Appl. Geophys.*, **131**(1–2), 197–210.

Matérn, B., 1960. *Spatial Variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations*, vol. 49, Statens Skogsforskningsintitut, Stockholm, Sweden.

Metropolis, N. & Ulam, S., 1949. The Monte Carlo method, *J. Am. Stat. Assoc.*, **44**(247), 335–341, doi: 10.1080/01621459.1949.10483310.

Montagner, J.-P., 1986. Regional three-dimensional structures using long-period surface waves, *Ann. Geophys.–Germany*, **4**(B3), 283–294.

Munk, W. H., 1955. High frequency spectrum of ocean waves, *J. Mar. Res.*, **14**(4), 302–314.

Neeser, F. D. & Massey, J. L., 1993. Proper complex random processes with applications to information theory, *IEEE T. Inform. Theory*, **39**(4), 1293–1302.

North, G. R., Wang, J. & Genton, M. G., 2011. Correlation models for temperature fields, *J. Climate*, **24**, 5850–5862, doi: 10.1175/2011JCLI4199.1.

Olhede, S. C., 2008. Localisation of geometric anisotropy, *IEEE T. Signal Process.*, **56**(5), 2133–2138, doi: 10.1109/TSP.2007.912894.

Olhede, S. C., Ramíres, D. & Schreier, P. J., 2014. Detecting directionality in random fields using the monogenic signal, *IEEE T. Inform. Theory*, **60**(10), 6491–6510, doi: 10.1109/TIT.2014.2342734.

Paciorek, C. J., 2007. Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP package, *J. Stat. Softw.*, **19**(2), nihpa22751.

Paciorek, C. J. & Schervish, M. J., 2006. Spatial modelling using a new class of nonstationary covariance functions, *Environmetrics*, **17**(5), 483–506, 10.1002/env.785.

Pawitan, Y., 2001. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford Univ. Press, New York.

Percival, D. B. & Walden, A. T., 1993. *Spectral Analysis for Physical Applications, Multitaper and Conventional Univariate Techniques*, Cambridge Univ. Press, New York.

Persson, B. N. J., 2006. Contact mechanics for randomly rough surfaces, *Surf. Sci. Rep.*, **61**, 201–227, doi: 10.1016/j.surfrep.2006.04.001.

Porcu, E., Bevilacqua, M., Schaback, R. & Oates, C. J., 2024. The Matérn model: A journey through statistics, numerical analysis and machine learning, *Stat. Sci.*, **39**(3), 469–492, doi: 10.1214/24–STS923.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P., 2007. *Numerical Recipes: The Art of Scientific Computing*, Cambridge Univ. Press, New York, 3rd edn.

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

Rasmussen, C. E. & Williams, C. K. I., 2006. *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, Mass.

Reed, I. S., Lee, P. C. & Truong, T.-K., 2002. Spectral representation of fractional Brownian motion in n dimensions and its properties, *IEEE T. Inform. Theory*, **41**(5), 1439–1451, doi: 10.1109/18.412687.

Reich, B. J., Bandyopadhyay, D. & Bondell, H. D., 2013. A nonparametric spatial model for periodontal data with nonrandom missingness, *J. Am. Stat. Assoc.*, **108**(503), 820–831, doi: 10.1080/01621459.2013.795487.

Rodríguez-Iturbe, I. & Mejía, J. M., 1974. The design of rainfall networks in time and space, *Water Resources Res.*, **10**(4), 713–728, doi: 10.1029/WR010i004p00713.

Rosenburg, M. A., Aharonson, O., Head, J. W., Kreslavsky, M. A., Mazarico, E., Neumann, G. A., Smith, D. E., Torrence, M. H. & Zuber, M. T., 2011. Global surface slopes and roughness of the Moon from the Lunar Orbiter Laser Altimeter, *J. Geophys. Res.*, **116**, E02001, doi: 10.1029/2010JE003716.

Sagy, A., Brodsky, E. E. & Axen, G. J., 2007. Evolution of fault-surface roughness with slip, *Geology*, **35**(3), 283–286, doi: 10.1130/G23235A.1.

Sandwell, D. T., Goff, J. A., Gevorgian, J., Harper, H., Kim, S.-S., Yu, Y., Tozer, B., Wessel, P. & Smith, W. H. F., 2022. Improved bathymetric prediction using geological information: SYNBATH, *Earth Space Sci.*, **9**(2), e2021EA002069.

Severini, T. A., 2001. *Likelihood Methods in Statistics*, Oxford Univ. Press, Oxford, UK.

Sharpton, V. L. & Head, J. W., 1985. Analysis of regional slope characteristics on Venus and Earth, *J. Geophys. Res.*, **90**(B5), 3733–3740, doi: 10.1029/JB090iB05p03733.

Shepard, M. K., Campbell, B. A., Bulmer, M. H., Farr, T. G., Gaddis, L. R. & Plaut, J. J., 2001. The roughness of natural terrain: A planetary and remote sensing perspective, *J. Geophys. Res.*, **106**(E12).

Simon, J. D., Simons, F. J., Irving, J. C. E., Wu, W., Obayashi, M., Yu, Y., Chen, Y. J., Sugioka, H. & Hello, Y. M., 2026. Hydroacoustic observations of the 15 January 2022 Hunga Tonga-Hunga Ha'apai eruption: The role of bathymetry along the path, *J. Geophys. Res.*, **131**, e2025JB032996, doi: 10.1029/2025JB032996.

Simons, F. J. & Olhede, S. C., 2013. Maximum-likelihood estimation of lithospheric flexural rigidity, initial-loading fraction and load correlation, under isotropy, *Geophys. J. Int.*, **193**(3), 1300–1342, doi: 10.1093/gji/ggt056.

Smith, W. H. F. & Sandwell, D. T., 1997. Global sea floor topography from satellite altimetry and ship depth soundings, *Science*, **277**(5334), 1956–1962, doi: 10.1126/science.277.5334.1956.

Song, T.-R. A. & Simons, M., 2003. Large trench-parallel gravity variations predict seismogenic behavior in subduction zones, *Nature*, **301**, 630–633, doi: 10.1126/science.1085557.

Stein, M. L., 1995. Fixed-domain asymptotics for spatial periodograms, *J. Am. Stat. Assoc.*, **90**(432), 1277–1288, doi: 10.1080/01621459.1995.10476632.

Stein, M. L., 1999. *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Series in Statistics, Springer, New York.

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

Stephenson, R., 1984. Flexural models of continental lithosphere based on the long-term erosional decay of topography, *Geophys. J. R. Astron. Soc.*, **77**, 385–413.

Sun, Y., Bowman, K. P., Genton, M. G. & Tokay, A., 2015. A Matérn model of the spatial covariance structure of point rain rates, *Stoch. Env. Res. Risk A.*, **29**, 411–416, doi: 10.1007/s00477–014–0923–2.

Sykulski, A. M., Olhede, S. C., Guillaumin, A. P., Lilly, J. M. & Early, J. J., 2019. The debiased Whittle likelihood, *Biometrika*, **106**, 251–266, doi: 10.1093/biomet/asy071.

Tarantola, A. & Nercessian, A., 1984. Three-dimensional inversion without blocks, *Geophys. J. R. Astron. Soc.*, **76**(2), 299–306, doi: 10.1111/j.1365–246X.1984.tb05047.x.

Valentine, A. P. & Davies, D. R., 2020. Global models from sparse data: A robust estimate of Earth's residual topography spectrum, *Geochem. Geophys. Geosys.*, **21**(8), e2020GC009240, doi: 10.1029/2020GC009240.

Vanmarcke, E., 1983. *Random Fields. Analysis and Synthesis*, MIT Press, Cambridge, Mass., 1st edn.

Vanmarcke, E., 2010. *Random Fields. Analysis and Synthesis*, World Scientific, Singapore, 2nd edn.

Vecchia, A. V., 1988. Estimation and model identification for continuous spatial processes, *J. R. Stat. Soc., Ser. B*, **50**(2), 297–312, doi: 10.1111/j.2517–6161.1988.tb01729.x.

Walden, A. T., McCoy, E. J. & Percival, D. B., 1994. The variance of multitaper spectrum estimates for real Gaussian processes, *IEEE T. Signal Process.*, **2**, 479–482.

Wang, K., Abdulah, S., Sun, Y. & Genton, M. G., 2023. Which parameterization of the Matérn covariance function?, *Spat. Stat.*, **58**, 100787, doi: 10.1016/j.spasta.2023.100787.

Watson, G. N., 1995. *A Treatise on the Theory of Bessel Functions*, Cambridge Univ. Press, Cambridge, UK, 2nd edn.

Whittaker, J. M., Müller, R. D., Roest, W. R., Wessel, P. & Smith, W. H. F., 2008. How supercontinents and superoceans affect seafloor roughness, *Nature*, **456**, 938–941, doi: 10.1038/nature07573.

Whittle, P., 1953. Estimation and information in stationary time series, *Arkiv Mat.*, **2**(23), 423–434.

Whittle, P., 1954. On stationary processes in the plane, *Biometrika*, **41**(3–4), 434–449, doi: 10.2307/2332724.

Wieczorek, M. A., 2015, The gravity and topography of the terrestrial planets, in *Treatise on Geophysics*, edited by T. Spohn, vol. 10, pp. 153–2193, doi: 10.1016/B978–0–444–53802–4.00169–X, Elsevier, Amsterdam, Neth., 2nd edn.

Wu, R.-S. & Aki, K., 1985. Elastic wave scattering by a random medium and the small-scale inhomogeneities in the lithosphere, *J. Geophys. Res.*, **90**(B12), 10261–10273, doi: 10.1029/JB090iB12p10261.

Wu, R.-S. & Flatté, S. M., 1990. Transmission fluctuations across an array and heterogeneities in the crust and upper mantle, *Pure Appl. Geophys.*, **132**(1–2), 175–196, doi: 10.1007/BF00874362.

Wunsch, C., 2022. Can oceanic flows be heard? Abyssal melodies, *J. Acoust. Soc. Am.*, **152**(4), 2160–2168, doi: 10.1121/10.0014603.

Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics, *J. Am. Stat. Assoc.*, **99**(465), 250–261, doi: 10.1198/016214504000000241.

Zhang, H. & Zimmerman, D. L., 2005. Towards reconciling two asymptotic frameworks in spatial statistics, *Biometrika*, **92**(4), 921–936, doi: 10.1093/biomet/92.4.921.

9 APPENDIX

The explicit forms of eq. (41), the logarithmic derivatives of the isotropic Matérn spectral density (12) in each of the three parameters $\theta \in \boldsymbol{\theta}$, which enter into the definition of the score (40) and the Fisher matrix (43) of the likelihood (39), are given in terms of the auxiliary variable

$$\mu = \left(\frac{4\nu}{\pi^2\rho^2}\right)\left(\frac{4\nu}{\pi^2\rho^2} + k^2\right)^{-1}, \tag{72}$$

by which we rewrite eqs A25–A27 Simons & Olhede (2013) as

$$m_{\sigma^2} = \frac{1}{\sigma^2}, \qquad m_\nu = \frac{\nu+1}{\nu} - \frac{\nu+1}{\nu}\mu + \ln\mu, \qquad \text{and} \qquad m_\rho = -2\frac{\nu}{\rho} + 2\frac{\nu+1}{\rho}\mu. \tag{73}$$

The nonvanishing derivatives of (73) necessary for the computation of the Hessian (42) are

$$\frac{\partial m_{\sigma^2}}{\partial\sigma^2} = -\frac{1}{\sigma^4}, \qquad \frac{\partial m_\nu}{\partial\nu} = \frac{1}{\nu}\left(\frac{\nu+1}{\nu} - \mu + \frac{\nu+1}{\nu}\mu^2\right), \qquad \frac{\partial m_\rho}{\partial\rho} = 2\frac{\nu}{\rho^2}\left(1 - 3\frac{\nu+1}{\nu}\mu + 2\frac{\nu+1}{\nu}\mu^2\right), \tag{74}$$

and the two identical cross-derivatives are

$$\frac{\partial m_\nu}{\partial\rho} = \frac{\partial m_\rho}{\partial\nu} = \frac{2}{\rho}\left(-1 + \frac{2\nu+1}{\nu}\mu - \frac{\nu+1}{\nu}\mu^2\right). \tag{75}$$

We list them for completeness only, since they are not needed for our analysis, though we used them for numerical checks.

Obtaining explicit forms for eq. (53), the logarithmic derivatives of the blurred spectral density (50) necessary to calculate the score (52) and the Fisher matrix (54) of the likelihood (48), requires derivatives of the isotropic Matérn spatial covariance (15) in the parameters $\theta \in \boldsymbol{\theta}$.

The first, in terms of the variance σ^2, is trivial,

$$\frac{\partial\mathcal{C}_{\boldsymbol{\theta}}(r)}{\partial\sigma^2} = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{2\nu^{\frac{1}{2}}}{\pi\rho}r\right)^\nu K_\nu\left(\frac{2\nu^{\frac{1}{2}}}{\pi\rho}r\right). \tag{76}$$

The second, in terms of smoothness ν, has a compact analytical form when $\nu = n \in \mathbb{Z}$, for which Abramowitz & Stegun (1965) eq. (9.6.45) provide the integer-order derivative of the modified Bessel function of the second kind. Using the shorthand $z = \left(2n^{1/2}r\right)/(\pi\rho)$, we have (see also Geoga et al. 2023),

$$\frac{\partial\mathcal{C}_{\boldsymbol{\theta}}(r)}{\partial n} = \sigma^2\frac{2^{1-n}}{\Gamma(n)}z^n K_n(z)\left(\frac{1}{2} + \frac{z}{2} + \ln\left(\frac{z}{2}\right) - \psi(n) + n!\left(\frac{z}{2}\right)^{-n}\sum_{m=0}^{n-1}\frac{(z/2)^m K_m(z)}{2(n-m)m!}\right), \quad \text{with} \quad \psi(n) = \frac{1}{\Gamma(n)}\frac{d\Gamma(n)}{dn}. \tag{77}$$

At large arguments and for non-integer ν we resort to a central-difference approximation that follows the guidance of Press et al. (2007) eq. (5.7.8) in selecting a ν-dependent step size to minimize truncation and round-off errors, that is $\epsilon^{1/3}\nu$, with ϵ the machine precision.

The third, with respect to range ρ, is found from the chain rule and Watson (1995) eq. (3.71.3), $zK'_\nu(z) + \nu K_\nu(z) = -zK_{\nu-1}(z)$,

$$\frac{\partial\mathcal{C}_{\boldsymbol{\theta}}(r)}{\partial\rho} = \frac{\sigma^2}{\rho}\frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{2\nu^{\frac{1}{2}}}{\pi\rho}r\right)^{\nu+1} K_{\nu-1}\left(\frac{2\nu^{\frac{1}{2}}}{\pi\rho}r\right). \tag{78}$$