

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

Frederik J. Simons^{1,2}, Olivia L. Walbert¹, Arthur P. Guillaumin³, Gabriel L. Eggers⁴, Kevin W. Lewis⁵, and Sofia C. Olhede⁶

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, 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

3 June 2025

SUMMARY

We present a spectral-domain maximum-likelihood procedure to invert for the structure of univariate random Gaussian fields parameterized using a Matérn covariance structure. Under isotropy, the three parameters are the process variance, smoothness, and range. We treat discretization and edge effects for finite regions in simulation, parameter estimation via Whittle likelihood maximization, and uncertainty quantification. As even the ‘best’ estimate may not be ‘good enough’: we test whether the ‘model’ itself warrants rejection. Our results are relevant for spatial interpolation or out-of-sample extension (e.g. kriging), machine learning and feature detection, on geological data. We present procedural details and focus on high-level results that have real-world implications for the study of Earth, planets, and moons.

Key words: Statistical analysis. Covariance estimation.

1 INTRODUCTION

What numbers, which statistical notions capture the “essence” of a patch of planetary topography? If it were a stationary, white, Gaussian process, simply, mean and variance of the data would be sufficient. However, computed over differing baselines, or at varying resolution, sample means and variances would fluctuate non-erratically, hence “whiteness” would prove immediately to be an untenable assumption. By reporting statistics over changing baselines (e.g., Sharpton & Head 1985; Aharonson et al. 1998; Rosenburg et al. 2011), we have effectively subscribed to the topography being given by a spatially *correlated* (stationary, Gaussian) process. Estimating the variance of the field (at a point), with little bias and with a reasonable estimation variance, requires knowledge of the covariance (between pairs of points). We are led to the estimation of (the parameters of) a spatial covariance function, or alternatively and equivalently, of a spectral density, from sampled data. The first (spatial estimation) is notoriously noisy, since it requires finding data pairs at increasing offsets. The second (spectral estimation), is notoriously affected by edge effects, Fourier-domain artifacts which lead to estimation bias, especially for multidimensional data sets.

Common choices for parameterized covariance functions of Gaussian random fields are exponential or squared-exponential (Gaussian), defined by a variance and a correlation range. Both are special cases of the Matérn class of covariance function (Guttorp & Gneiting 2006), with a specific smoothness or mean-squared differentiability, itself an inversion parameter of interest (Goff & Jordan 1989a,b). In this paper we present a spectral-domain “Whittle” maximum-likelihood procedure that estimates the variance, smoothness, and range of an isotropic Gaussian point process, from sampled data. We show how to obtain unbiased estimates for these size and shape parameters when the region under study is not rectangular or circular, nor regularly sampled, and we calculated their estimation covariance, such that the results from differently sized and sampled patches can be compared robustly.

Our results have widespread implications for the study of planetary topography in the solar system, 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 histories for the patches under consideration. Our results should also be relevant for whomever needed to perform spatial interpolation or out-of-sample extension (e.g. via kriging) on geological data (Stein 1999). They also carry implications for machine learning and feature detection (Rasmussen & Williams 2006). We will present procedural details but also focus on high-level results that have real-world implications for the study of Venus, Earth, other planets, and moons. We illustrate our methodology on geologically mapped patches of Venus (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 Wiczorek (2015). To make the jump from geology and geophysics to statistics, in particular 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

Planetary topography $\mathcal{H}(\mathbf{x})$ is considered to be a zero-mean, finite-variance, stationary, two-dimensional random field. Under what is known as the Cramér (1942) representation, there exists a continuous spectral increment process, $d\mathcal{H}(\mathbf{k})$, according to which the spatial field

$$\mathcal{H}(\mathbf{x}) = \iint e^{i\mathbf{k}\cdot\mathbf{x}} d\mathcal{H}(\mathbf{k}). \quad (1)$$

The integration is over the space containing all wave vectors \mathbf{k} . In the case of strict band-limitation 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 realizations,

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

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

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

where $\delta(\mathbf{k}, \mathbf{k}')$ is the Dirac delta function. When eqs (1)–(3) hold, the spatial auto-covariance, $\mathcal{C}(\mathbf{x}, \mathbf{x}')$, betrays 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

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

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,

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

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

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

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

2.1.2 Isotropy

Under isotropy, $\mathcal{S}(\mathbf{k}) = \mathcal{S}(k)$, 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,

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

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

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

Introducing the distance variable r , we rewrite eq. (7) as

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

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

Isotropy remains a restrictive—but testable—assumption, which we will be relaxing in future work.

2.1.3 Matérnity

We further specify planetary topography as a member of the Matérn class (Stein 1999), which is very general and widely applicable (Rasmussen & Williams 2006). The isotropic d -dimensional Matérn spectral density $\mathcal{S}_\theta^d(k)$ assumes the parameterized form

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

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

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

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

$$\boldsymbol{\theta} = [\sigma^2 \ \nu \ \rho]^T. \quad (13)$$

The ‘variance’, σ^2 , indeed satisfies eq. (9) upon substitution with eq. (12). At short wavelengths, when k is large, the spectrum $\mathcal{S}_\theta(k)$ 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 k small, is controlled by the combined effect of σ^2 and ρ . The fluctuation scale

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

The isotropic Matérn spatial covariance $\mathcal{C}_\theta(r)$ is unlike its spectral counterpart (11) in requiring no dimensional specification,

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

with K_ν the modified Bessel function of the second kind. The asymptotic behavior $K_\nu(z) \rightarrow \Gamma(\nu) (z/2)^{-\nu}/2$ for small z , verifies that $\mathcal{C}_\theta(0) = \sigma^2$ as in eq. (9). For low values of ν , furthermore, $\mathcal{C}_\theta(\pi\rho) \approx \sigma^2/3$. 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

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

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

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

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

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

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 extreme generality of the Matérn class is appreciated by evaluating the correlation functions for special values of ν . Notably, when $\nu = 1/2$, the correlation function decays exponentially, and when $\nu \rightarrow \infty$, 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 (Guttorp & Gneiting 2006). Despite all of its generality, it is important to not identify our conceptual point of departure with fractal, scale-invariant, self-affine, or self-similar behavior (see, e.g. Mareschal 1989; Herzfeld et al. 1995), about which we make no claims.

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 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 \mathbf{x} maps out a rectangular $M \times N$ grid given by

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

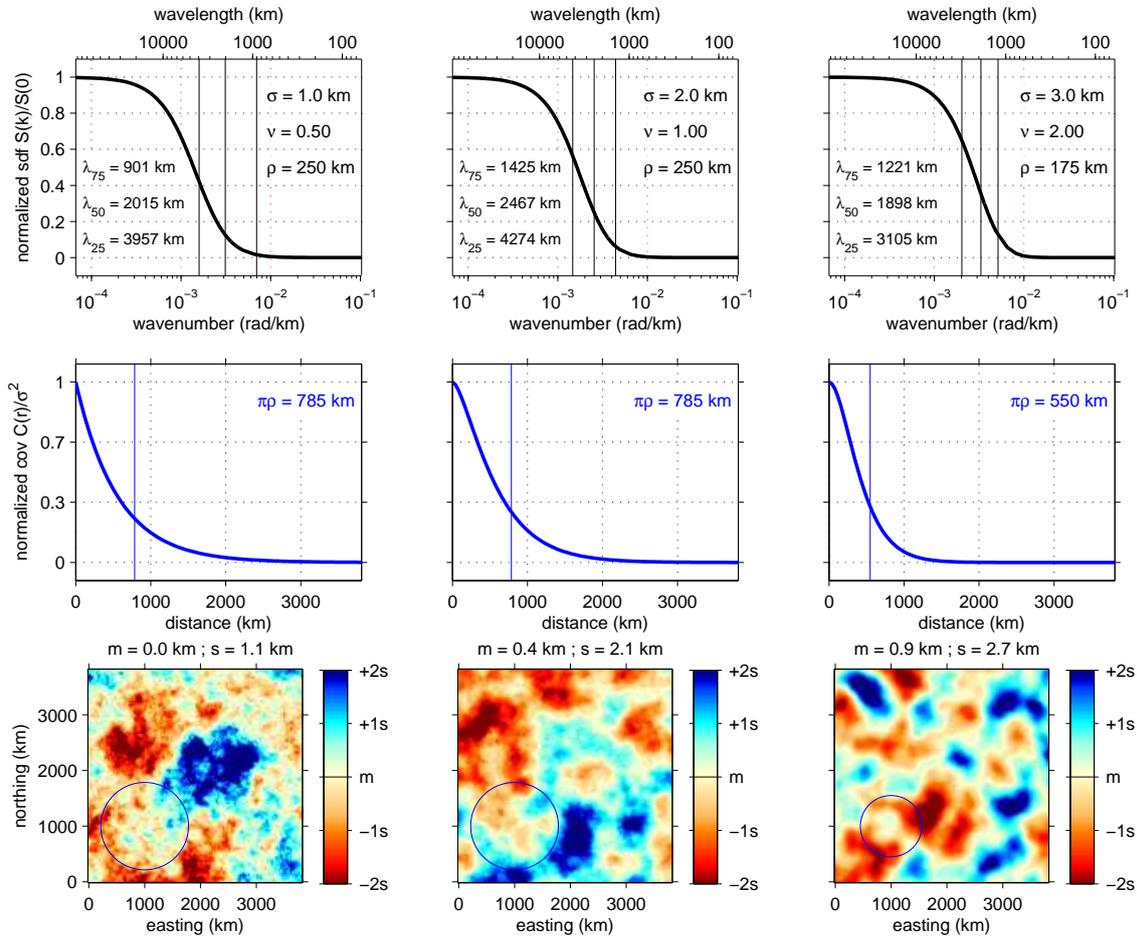


Figure 1. Random fields generated from stationary isotropic Matérn models with variances σ^2 , differentiabilitys ν , and correlation lengths ρ , as indicated. (*Top:*) Normalized spectral densities, $\mathcal{S}_\theta(k)/\mathcal{S}_\theta(0)$, from eq. (12). The vertical black lines identify the wavenumbers 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, $\mathcal{C}_\theta(r)/\sigma^2$, 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.

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

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

Sampled in spectral space, the wave vectors are, with m and n as in eq. (19),

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

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

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

2.2.2 Blurring

Obtaining space-domain realizations from a population of random fields specified by a certain spectral density, e.g., 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 a true quantity $\mathcal{S}_\theta(\mathbf{k})$ and showed no correlations between wavenumbers. Simons & Olhede (2013) showed (for multivariate fields) that the covariance of a finite set of gridded Fourier coefficients can at best offer a blurred and correlated version of the true spectral variance (their eq. 9), and they

simulated fields by incorporating the blurring but ignoring the correlation (their eq. 83). They carried out the blurring computationally, and approximately, via grid refinement, convolution, and subsampling (their eq. 89).

Guillaumin et al. (2022) (their Lemmata 1 and 2) showed how to incorporate the spectral blurring effect of applying arbitrary data windows, including irregular boundaries and incomplete sampling, exactly, and at a much reduced computational cost. We rewrite the discrete Fourier transform in eqs (19) and (20) to incorporate an arbitrary unit-normalized data window, $w(\mathbf{x})$,

$$H(\mathbf{k}) \equiv \frac{1}{2\pi} \left(\frac{\Delta x \Delta y}{MN} \right)^{\frac{1}{2}} \sum_{\mathbf{x}} w(\mathbf{x}) \mathcal{H}(\mathbf{x}) e^{-i\mathbf{k} \cdot \mathbf{x}}. \quad (23)$$

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

$$\text{var} \{H(\mathbf{k})\} = \frac{1}{(2\pi)^2} \left(\frac{\Delta x \Delta y}{MN} \right) \sum_{\mathbf{x}} \sum_{\mathbf{x}'} w(\mathbf{x}) w(\mathbf{x}') \mathcal{C}_{\theta}(\mathbf{x} - \mathbf{x}') e^{-i\mathbf{k} \cdot (\mathbf{x} - \mathbf{x}')}, \quad (24)$$

$$= \frac{1}{(2\pi)^2} \left(\frac{\Delta x \Delta y}{MN} \right) \sum_{\mathbf{y}} \left(\sum_{\mathbf{x} - \mathbf{y} \cap \mathbf{x}} w(\mathbf{x}) w(\mathbf{x} - \mathbf{y}) \right) \mathcal{C}_{\theta}(\mathbf{y}) e^{-i\mathbf{k} \cdot \mathbf{y}}, \quad (25)$$

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

$$\mathbf{y} = \{(m' \Delta x, n' \Delta y)\}, \quad \text{with the mirrored index sets} \quad m' = -M + 1, \dots, M - 1 \quad \text{and} \quad n' = -N + 1, \dots, N - 1, \quad (26)$$

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

$$W(\mathbf{y}) = \sum_{\max(\mathbf{x} - \mathbf{y}, |\mathbf{y}|)} w(\mathbf{x}) w(\mathbf{x} - |\mathbf{y}|) = \sum_{\min(\mathbf{x} - \mathbf{y}, |\mathbf{y}|)} w(\mathbf{x}) w(\mathbf{x} + |\mathbf{y}|). \quad (27)$$

Reviewing what this implies for the variance of the Fourier coefficient of a windowed and sampled field $\mathcal{H}(\mathbf{x})$, we rewrite eqs (24)–(25) as

$$\text{var} \{H(\mathbf{k})\} = \frac{1}{(2\pi)^2} \left(\frac{\Delta x \Delta y}{MN} \right) \sum_{\mathbf{y}} W(\mathbf{y}) \mathcal{C}_{\theta}(\mathbf{y}) e^{-i\mathbf{k} \cdot \mathbf{y}} = \bar{\mathcal{S}}_{\theta}(\mathbf{k}), \quad (28)$$

the exact version, valid for arbitrary data windows, of what Simons & Olhede (2013) (in their eq. 89) implemented approximately and only for rectangular windows, by discrete convolution of the underlying spectral density $\mathcal{S}_{\theta}(\mathbf{k})$ with the Fejér kernel. In that special case of a unitary window function, that is, for complete observations on a rectangular grid, eq. (27) evaluates to

$$W(\mathbf{y}) = (M - |m'|)(N - |n'|). \quad (29)$$

Used as a basis for simulation, generating Fourier coefficients $H(\mathbf{k})$ from the Cholesky decomposition of $\bar{\mathcal{S}}_{\theta}(\mathbf{k})$, ignores wavenumber correlation effects. To make Fig. 1, we generated four times as many spatial samples as needed, retaining only the central portion for analysis, to avoid wrap-around correlations. Constructing spatial patches via eq. (22), on the space grid (19), the variance 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 is stationary,

$$\langle \mathcal{H}(\mathbf{x}) \mathcal{H}(\mathbf{x}') \rangle = \mathcal{C}_{\theta}(\mathbf{x} - \mathbf{x}'). \quad (30)$$

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

$$\text{var} \{ \mathcal{H}(\mathbf{x}) \} \approx \frac{(2\pi)^2 \sum_{\mathbf{k}} \mathcal{S}_{\theta}(\mathbf{k})}{(MN \Delta x \Delta y)} \approx \sigma^2. \quad (31)$$

Eq. (28) shows that the *expected periodogram* of the data can be obtained via Fourier transformation of the autocovariance sequence of the sampling window. Fig. 2 shows this equivalence. From a sequence of realizations, we show, in the top row from left to right, 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 $\bar{\mathcal{S}}_{\theta}(\mathbf{k})$, for the parameter set θ 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 the bottom row, we show a unit normalized square taper, the ratio of the average periodogram to its expectation, $\text{mean} |H(\mathbf{k})|^2 / \bar{\mathcal{S}}_{\theta}(\mathbf{k})$, 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 $\bar{\mathcal{S}}_{\theta}(\mathbf{k})$ shown directly above.

2.2.3 Simulation

Simons & Olhede (2013) (their eq. 9) showed 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 the off-diagonal terms in $\langle H(\mathbf{k}) H^*(\mathbf{k}') \rangle$ decay rapidly enough in most cases for us to ignore them as the basis for the simulations that we just showed, which took $\langle H(\mathbf{k}) H^*(\mathbf{k}') \rangle \approx \bar{\mathcal{S}}_{\theta}(\mathbf{k})$ as their point of departure, whether calculated on the interior domain of a doubled grid, using grid refinement, discrete convolution, and subsampling to approximate $\bar{\mathcal{S}}_{\theta}$, or exactly, via eq. (28). Second, we will show empirically that we are able to ignore them when designing the debiased-Whittle likelihood to perform parameter estimates

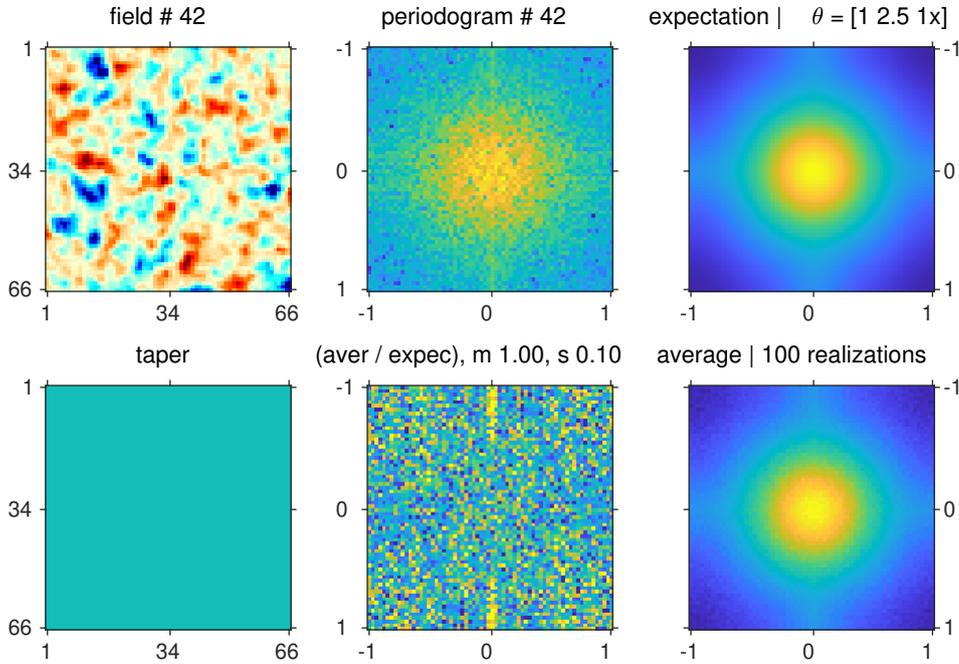


Figure 2. The Fejér blurred spectral density $\bar{S}_\theta(\mathbf{k})$ approximates the expectation of the periodogram, $|H(\mathbf{k})|^2$, of gridded and (unit) tapered data generated from the population density $S_\theta(\mathbf{k})$. (*Top row*) A single realization, $\mathcal{H}(\mathbf{x})$, its modified periodogram $|H(\mathbf{k})|^2$, and the blurred spectrum $\bar{S}_\theta(\mathbf{k})$. (*Bottom row*) The unit taper, the ratio of the average periodogram to the blurred spectral density, and the average periodogram, over 100 realizations.

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.

As to data simulation, we are now able to take wavenumber correlation into account also by switching to space-domain methods that use the spatial covariance, eq. (15), as their point of departure, via circulant embedding of the covariance matrix (Kroese & Botev 2015).

3 WHITHER WHITTLE ?

In this paper we develop a maximum-likelihood procedure that takes gridded input ‘topographies’ and estimates the three-element sets θ , see eq. (13), that contain the parameters of the isotropic Matérn spectral densities $S_\theta(k)$ or spatial covariances $C_\theta(r)$ 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

$$s^2 = \frac{1}{MN} \sum_{\mathbf{x}} \mathcal{H}^2(\mathbf{x}) - \frac{1}{(MN)^2} \left(\sum_{\mathbf{x}} \mathcal{H}(\mathbf{x}) \right)^2. \quad (32)$$

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 (30) and (31), we find from eq. (32) that in expectation, approximately,

$$\langle s^2 \rangle \approx \frac{1}{MN} \sum_{\mathbf{x}} \frac{(2\pi)^2 \sum_{\mathbf{k}} S_\theta(\mathbf{k})}{(MN\Delta x\Delta y)} - \frac{1}{(MN)^2} \sum_{\mathbf{x}} \sum_{\mathbf{x}'} C(\mathbf{x} - \mathbf{x}') \approx \sigma^2 - \frac{1}{(MN)^2} \sum_{\mathbf{x}} \sum_{\mathbf{x}'} C(\mathbf{x} - \mathbf{x}') \quad (33)$$

$$\approx \sigma^2 - \frac{(2\pi)^2 \bar{S}_\theta(0)}{MN\Delta x\Delta y}. \quad (34)$$

Eq. (33) applies quite generally, to stationary processes with spectral density $S_\theta(\mathbf{k})$ or covariance function $C_\theta(\mathbf{x})$. The second term in eq. (34) can be interpreted as discretizing eq. (7). The appearance of the *blurred* spectral density, $\bar{S}_\theta(0)$, 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, $C_\theta(\mathbf{x} - \mathbf{x}') = \sigma^2 \delta(\mathbf{x}, \mathbf{x}')$, does eq. (33) reduce to the iid bias expression (Bendat & Piersol 2000)

$$\langle s^2 \rangle \approx \sigma^2 \left(1 - \frac{1}{MN} \right). \quad (35)$$

If the spatial grid is fine enough, and the full behavior of the spatial covariance $C(\mathbf{x})$ is being accurately captured by the sampling density,

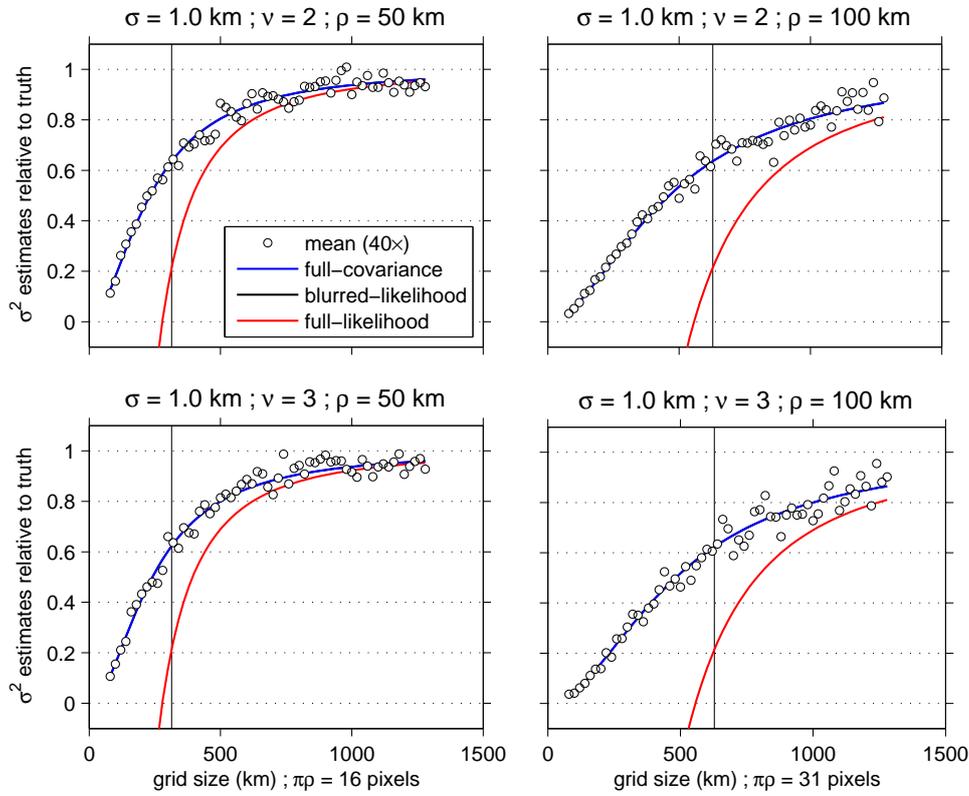


Figure 3. The sample variance s^2 systematically underestimates the true process variance σ^2 . It is negatively biased by the presence of spatial correlation embodied by the Matérn parameters ν and ρ , listed in the titles. Black open circles (‘mean’) are averages of the sample variances s^2 for data 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. (33), 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. (34). Solid red lines (‘full-likelihood’) are from eq. (36). 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.

relatively to the scale length of correlation, $\mathcal{S}_\theta(0)$ can again be substituted for $\bar{\mathcal{S}}_\theta(0)$ in eq. (34). In that case, using eq. (14) yields the form applicable to the isotropic Matérn density, namely

$$\langle s^2 \rangle \approx \sigma^2 \left(1 - \frac{\pi(\pi\rho)^2}{MN\Delta x\Delta y} \right). \quad (36)$$

While this last approximation is usually too crude for bias calculations, eq. (36) 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, and the more negatively biased the naive estimator eq. (32) 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, as developed in subsequent sections.

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.

It would take us too far, and down a dead-end street, to write down the expressions for the *variance* of the naive variance estimator s^2 . Rather, we will illustrate its behavior on the basis of another suite of numerical experiments. Fig. 4 (left panels) reveals that the estimation variance is generally high (relatively speaking), and decaying too slowly (for our taste) with increasing grid size.

In comparison, the 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 decorrelation 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, suggestively 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.

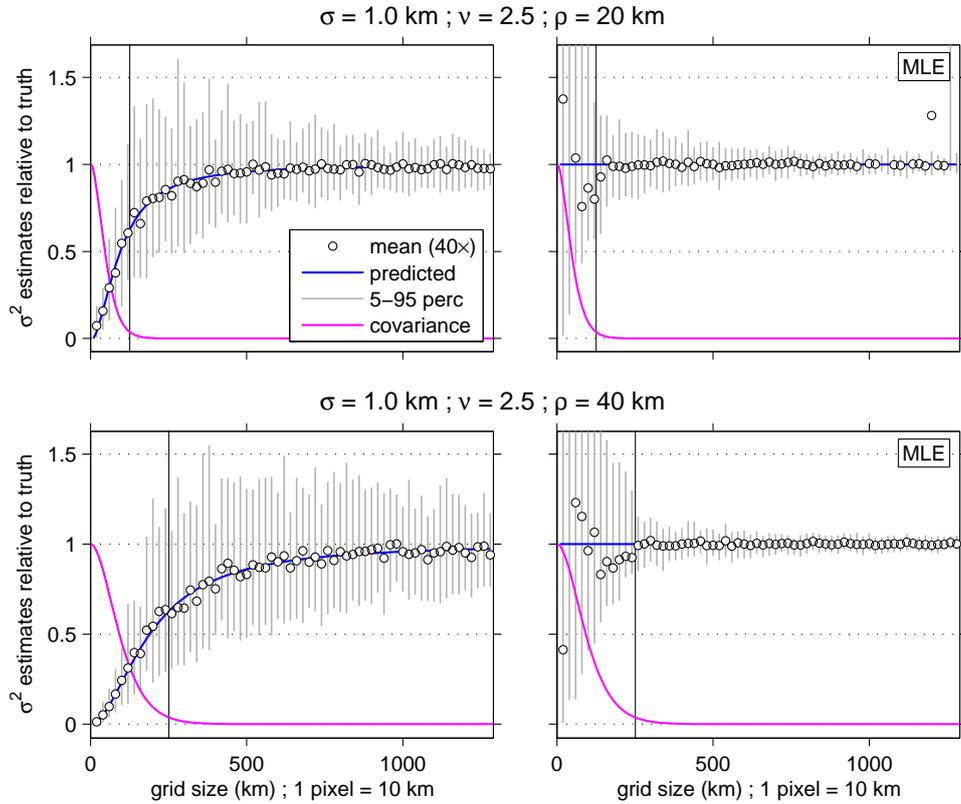


Figure 4. The sample variance s^2 is a biased, inconsistent, and inefficient estimator for the true process variance σ^2 . The maximum-likelihood estimator is asymptotically unbiased, consistent and efficient. Conducting 40 lattice simulations on differently sized data patches, with Matérn parameters (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 grey 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. (33), 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 ANALYSIS

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 for the notions developed in this section, though modern texts such as Pawitan (2001) and Severini (2001) are equally recommended. Our main device is the frequency-domain Whittle (1953, 1954) likelihood, blurred to acknowledge lattice effects.

4.1 Infinite sample size

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

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{MN} \sum_{\mathbf{k}} [\ln \mathcal{S}_\theta(k) + \mathcal{S}_\theta^{-1}(k) |H(\mathbf{k})|^2]. \quad (37)$$

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

$$\gamma_\theta(\boldsymbol{\theta}) = -\frac{1}{MN} \sum_{\mathbf{k}} m_\theta(k) [1 - \mathcal{S}_\theta^{-1}(k) |H(\mathbf{k})|^2], \quad (38)$$

with the required expressions

$$m_\theta(k) = \mathcal{S}_\theta^{-1}(k) \frac{\partial \mathcal{S}_\theta(k)}{\partial \theta} \quad (39)$$

easily obtained via differentiation of eq. (12), and listed as eqs (65)–(67) in the Appendix.

Its second derivatives are the elements of the Hessian matrix $\mathbf{F}(\boldsymbol{\theta})$,

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

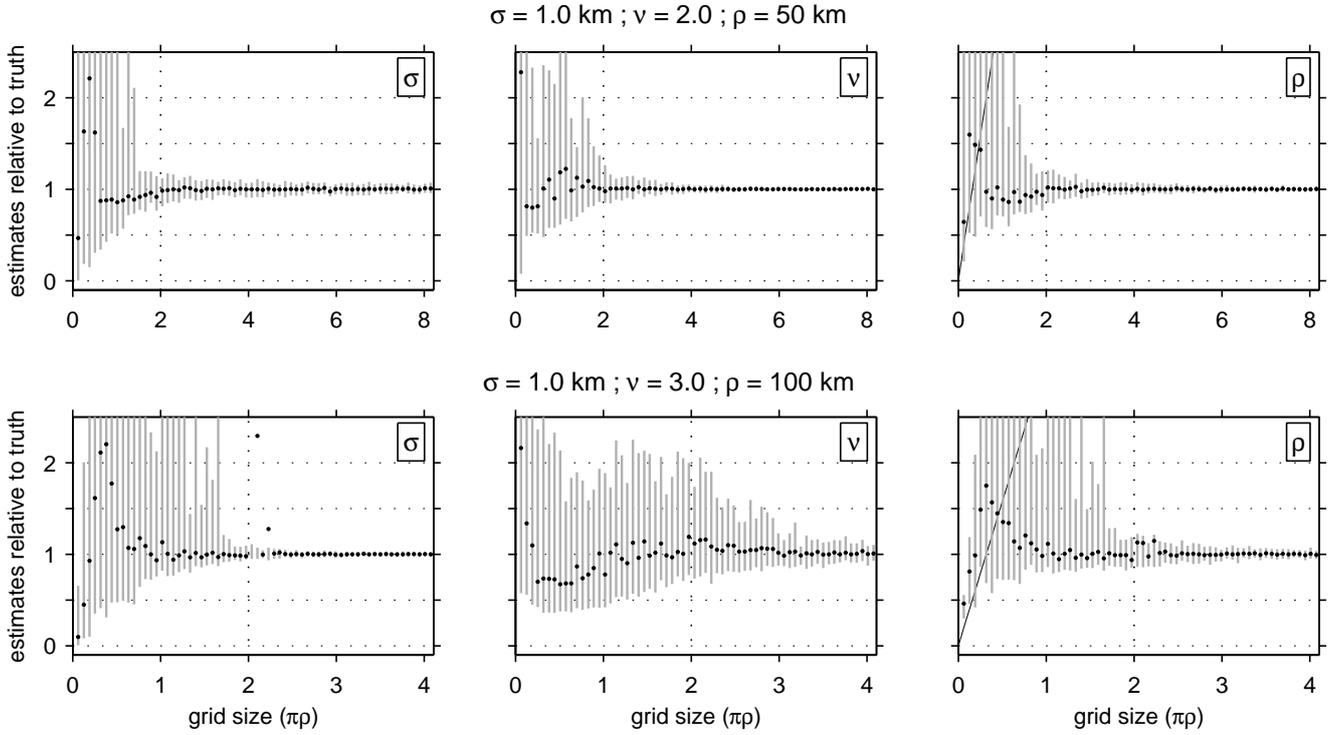


Figure 5. Behavior of the maximum-likelihood estimators of the Matérn parameters (σ^2, ν, ρ) for the two sets of true values listed in the titles, over 40 lattice simulations conducted on square data fields composed of up to $M = N = 128$ pixels of size $\Delta x = \Delta y = 10$ km. Grey bars cover the 5th through 95th percentiles of all of the estimates in each set. 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 40 estimates beyond that size. With growing field size, the estimates reveal themselves to be unbiased with very low variance. The slanted line in the rightmost panels corresponds to the data patch size.

with the nonvanishing derivatives $\partial_{\theta} m_{\theta'}(k)$ given as eqs (68)–(71) in the Appendix.

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

$$\mathcal{F}_{\theta\theta'}(\boldsymbol{\theta}) = \frac{1}{MN} \sum_{\mathbf{k}} m_{\theta}(k) m_{\theta'}(k). \quad (41)$$

The inverse of the Fisher matrix is the information matrix, $\mathcal{F}^{-1} = \mathcal{J}$. 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 (23) implies the distribution

$$\sqrt{MN}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim \mathcal{N}(\mathbf{0}, \mathcal{F}^{-1}(\boldsymbol{\theta}_0)) = \mathcal{N}(\mathbf{0}, \mathcal{J}(\boldsymbol{\theta}_0)), \quad (42)$$

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:

$$\hat{\theta} - z_{\beta/2} \frac{\mathcal{J}_{\theta\theta}^{1/2}(\hat{\boldsymbol{\theta}})}{\sqrt{MN}} \leq \theta_0 \leq \hat{\theta} + z_{\beta/2} \frac{\mathcal{J}_{\theta\theta}^{1/2}(\hat{\boldsymbol{\theta}})}{\sqrt{MN}}. \quad (43)$$

The relations in this section are theoretical quantities derived by Simons & Olhede (2013) that, strictly speaking, apply only to the ‘population’ case. As mentioned before, however, in the ‘sample’ case of discretized, windowed likelihood analysis, we must replace the Matérn spectral density $\mathcal{S}_{\boldsymbol{\theta}}(k)$ 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 for the uncertainty estimates of the parameters.

4.2 Sampled data, heuristics

The blurred likelihood $\bar{\mathcal{L}}(\boldsymbol{\theta})$ of observing the data $\mathcal{H}(\mathbf{x})$ under the spectral model (12)–(13) parameterized by $\boldsymbol{\theta}$ is given in terms of the Fourier coefficients of the gridded and windowed data, $H(\mathbf{k})$ in eq. (23), and of the blurred isotropic spectral density $\bar{\mathcal{S}}_{\boldsymbol{\theta}}(k)$ of eq. (28), summed over all wavenumbers in the grid (21), by the relation

$$\bar{\mathcal{L}}(\boldsymbol{\theta}) = -\frac{1}{MN} \sum_{\mathbf{k}} [\ln \bar{\mathcal{S}}_{\boldsymbol{\theta}}(k) + \bar{\mathcal{S}}_{\boldsymbol{\theta}}^{-1}(k) |H(\mathbf{k})|^2]. \quad (44)$$

Compare eq. (44) with eq. (37): the only difference is the acknowledgement of the spectral blurring effect of sampled data. Eq. (44) is the quantity that we maximize, under positivity constraints, for the parameter vector θ , thereby defining the maximum-likelihood estimate $\hat{\theta}$ to be the maximizer of the ‘score’ vector $\bar{\gamma}(\theta)$ of numerical first derivatives of the blurred likelihood, which solves

$$\bar{\gamma}(\hat{\theta}) = \mathbf{0}. \quad (45)$$

Satisfying eq. (45) to find $\hat{\theta}$, 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 discussed in Section 2.2.2. Without entering into the details for now, Fig. 5 shows the results of a suite of experiments conducted to illustrate the performance of a 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 $2\pi\rho$ onwards, 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 therefore 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 $\bar{S}_\theta(k)$ for $S_\theta(k)$ in eqs (38) and (40), 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 $S_\theta(k)$. Replacing $S_\theta(k)$ by $\bar{S}_\theta(k)$ in eq. (38) 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. (44), solving eq. (45) without being given explicit expressions for the score and the Hessian, but 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 Appendix A8 of Simons & Olhede 2013) tempted them into concluding that eq. (43) could be used to construct confidence intervals for the solutions of eq. (45) 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—and 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. (41), 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. (44) that rely on eq. (43) 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. 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 Sampled data, full theory

For sampled data the likelihood comprises the blurred spectral density \bar{S}_θ , and the squared windowed Fourier coefficients of the data $\mathcal{H}(\mathbf{x})$,

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

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

$$H(\mathbf{k}) \equiv \frac{1}{2\pi} \left(\frac{\Delta x \Delta y}{MN} \right)^{\frac{1}{2}} \sum_{\mathbf{x}} w(\mathbf{x}) \mathcal{H}(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}, \quad (47)$$

and the Matérn spectral density S_θ whose parameters we aim to recover (see eq. 12) is exactly blurred to account for finite-sample effects via the intermediary of the Matérn covariance \mathcal{C}_θ (see eq. 15), as in eq. (28),

$$\bar{S}_\theta(\mathbf{k}) = \frac{1}{(2\pi)^2} \left(\frac{\Delta x \Delta y}{MN} \right) \sum_{\mathbf{y}} W(\mathbf{y}) \mathcal{C}_\theta(\mathbf{y}) e^{-i\mathbf{k}\cdot\mathbf{y}}, \quad (48)$$

and with W the autocorrelation of the sampling window, that is,

$$W(\mathbf{y}) = \sum_{\mathbf{x}} w(\mathbf{x})w(\mathbf{x} + \mathbf{y}). \quad (49)$$

The solution $\hat{\theta}$ is found by maximization of eq. (46), requiring the vanishing of the score $\nabla \bar{\mathcal{L}}(\theta) = \bar{\gamma}(\theta)$, whose elements are given by

$$\bar{\gamma}_{\theta}(\theta) = -\frac{1}{MN} \sum_{\mathbf{k}} \bar{m}_{\theta}(k) [1 - \bar{S}_{\theta}^{-1}(k) |H(\mathbf{k})|^2], \quad (50)$$

with the blurred equivalents to eq. (39) again obtained exactly as

$$\bar{m}_{\theta} = \frac{\bar{S}_{\theta}^{-1}}{(2\pi)^2} \left(\frac{\Delta x \Delta y}{MN} \right) \sum_{\mathbf{y}} W(\mathbf{y}) \frac{\partial \mathcal{C}_{\theta}(\mathbf{y})}{\partial \theta} e^{-i\mathbf{k} \cdot \mathbf{y}}, \quad (51)$$

in which several of the derivatives in the parameters θ of the spatial covariance $\partial_{\theta} \mathcal{C}_{\theta}$ are analytically available, as we will write elsewhere.

The elements of the Fisher matrix $\bar{\mathcal{F}}(\theta)$ are now given by

$$\bar{\mathcal{F}}_{\theta\theta'}(\theta) = \frac{1}{MN} \sum_{\mathbf{k}} \bar{m}_{\theta}(k) \bar{m}_{\theta'}(k). \quad (52)$$

As Simons & Olhede (2013) (their eq. 138), but now following Guillaumin et al. (2022) (their eq. 36), the parameter estimation variance, $\text{cov}(\hat{\theta}) \approx \bar{\mathcal{F}}^{-1}(\theta_0) \text{cov}\{\bar{\gamma}(\theta_0)\} \bar{\mathcal{F}}^{-1}(\theta_0)$,

requires the additional calculation of the covariance of the score *without* neglecting the correlation between wavenumbers,

$$\text{cov}\{\gamma_{\theta}, \gamma_{\theta'}\} = \frac{1}{(MN)^2} \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \bar{m}_{\theta}(\mathbf{k}) \frac{\text{cov}\{|H(\mathbf{k})|^2, |H(\mathbf{k}')|^2\}}{\bar{S}(\mathbf{k})\bar{S}(\mathbf{k}')} \bar{m}_{\theta}(\mathbf{k}'), \quad (54)$$

see Guillaumin et al. (2022) (their eq. 37), and compare Simons & Olhede (2013) (their eq. A54), which implies that we require the covariance of the windowed periodogram, which can be written under standard theory as (Simons & Olhede 2013, their eq. A57)

$$\text{cov}\{|H(\mathbf{k})|^2, |H(\mathbf{k}')|^2\} = |\text{cov}\{H(\mathbf{k}), H^*(\mathbf{k}')\}|^2 + |\text{cov}\{H(\mathbf{k}), H(\mathbf{k}')\}|^2. \quad (55)$$

Eq. (54) can be approximated by simulation, but eq. (55) can be calculated exactly using one of two methods. Ultimately, this yields three different methods to obtain the desired estimation variance in eq. (54). Three/four if we mention but discard the Monte Carlo approach of Guillaumin et al. (2022)

4.3.1 Derivatives of the spatial covariance

Brief. Most of it in Paper II.

4.3.2 Exact method 1?

This probably *does* belong in this paper.

4.3.3 Exact method 2?

This probably *does* belong in this paper.

4.4 Numerical Examples

Illustrate for square grids in Paper I? And then France, anti France, and random, in Paper II? So Paper II starts from Section 4.3, then talks about \mathcal{S} and \mathcal{C} as Fourier transforms, with the special cases table, and there is parallellism from this paper's Figure 2, and these, and then simulations, and then we can go on with Venus in Paper III. I think that is the best approach.

Fig. 7 is a representative illustration of this type of behavior. We conducted hundreds of inversions for independent simulations, and studied the distribution of the estimates and investigated how well the theoretical expression for their covariance approximated the observed regime. Invariably, our estimates were 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 standardized probability density function of the values recovered in this experiment of sample size 64×64 . The abscissas were truncated to within ± 3 of the empirical standard deviation; the percentage of the values captured by this truncation is listed in the top left of each graph. The ratio of the empirical to theoretical standard deviation is shown listed as s/σ for each of the parameters. The bottom row shows the quantile-quantile plots of the empirical (ordinate) versus the theoretical (abscissa) 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.

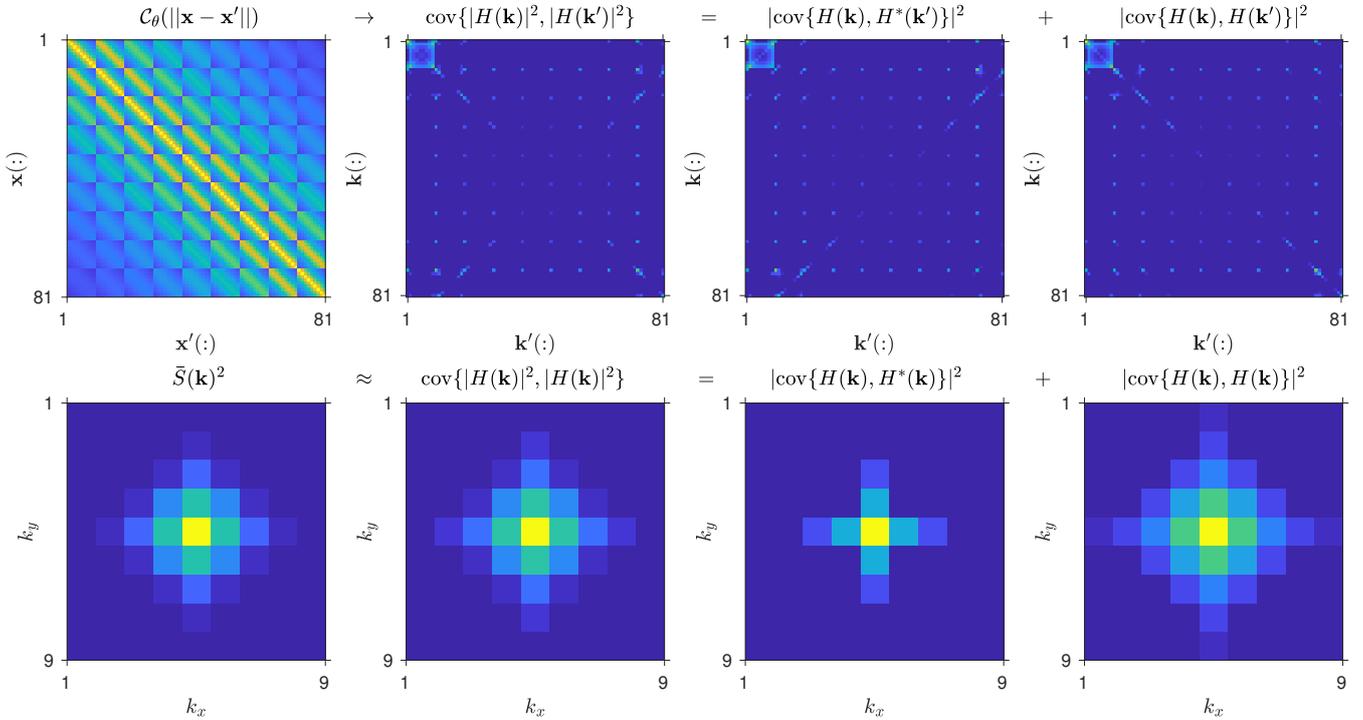


Figure 6. bluybla.

Fig. 8 is discussed. These are cross plots of the recovered parameters with their summaries and with the predicted covariances closely matching. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur condimentum venenatis magna. Ut vulputate urna vitae luctus sollicitudin. Etiam tempus nisl at mi imperdiet interdum. Pellentesque nec urna turpis. Etiam ultricies consectetur mauris hendrerit tristique.

Fig. 9 is discussed. These are normalized covariance plots of the recovered parameters with their predicted values from our complete theory. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur condimentum venenatis magna. Ut vulputate urna vitae luctus sollicitudin. Etiam tempus nisl at mi imperdiet interdum. Pellentesque nec urna turpis. Etiam ultricies consectetur mauris hendrerit tristique.

4.5 Analysis of residuals

The terms $\bar{S}_\theta^{-1}(k)|H(\mathbf{k})|^2$ and $S_\theta^{-1}(k)|H(\mathbf{k})|^2$ that have appeared above in the expressions for the likelihoods (44), (37) and their derivatives 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 have previously noted that if the Gaussian model fits, the expectation of this quadratic is $\langle \bar{S}_\theta^{-1}(k)|H(\mathbf{k})|^2 \rangle = 1$. Here we follow Simons & Olhede (2013) to maintain that twice this quantity should be a chi-squared random variable with two degrees of freedom,

$$X_\theta(\mathbf{k}) = \bar{S}_\theta^{-1}(k)|H(\mathbf{k})|^2 \sim \chi_2^2/2. \quad (56)$$

Equipped with this knowledge we can examine how closely the ratios $X_\theta(\mathbf{k})$, i.e., the ‘residuals’, follow the distribution (56), 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. (56) in fact holds can be carried out visually, by inspection of the overlay of the histograms of $2X_\theta$ across all wave vectors with the probability density function χ_2^2 , and by making ‘quantile-quantile’ plots of the ranked values of $2X_\theta$ versus the inverse cumulative density function of χ_2^2 evaluated at their corresponding fractional ranks. Moreover, the two-dimensional map of $X_\theta(\mathbf{k})$ should show no residual structure, and will contain information on possible wavenumber ranges or specific directions in which the data might be over- or under-fit. All three such representations of model quality must be thoroughly scrutinized.

Fig 10 enlightens us in this regard. The top three panels 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, and whose residuals $X_\theta(\mathbf{k})$ showed 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_θ are listed above the first panel. Per eq. (56), both are expected to be one. Above the second panel are the test statistic s_X^2 ,

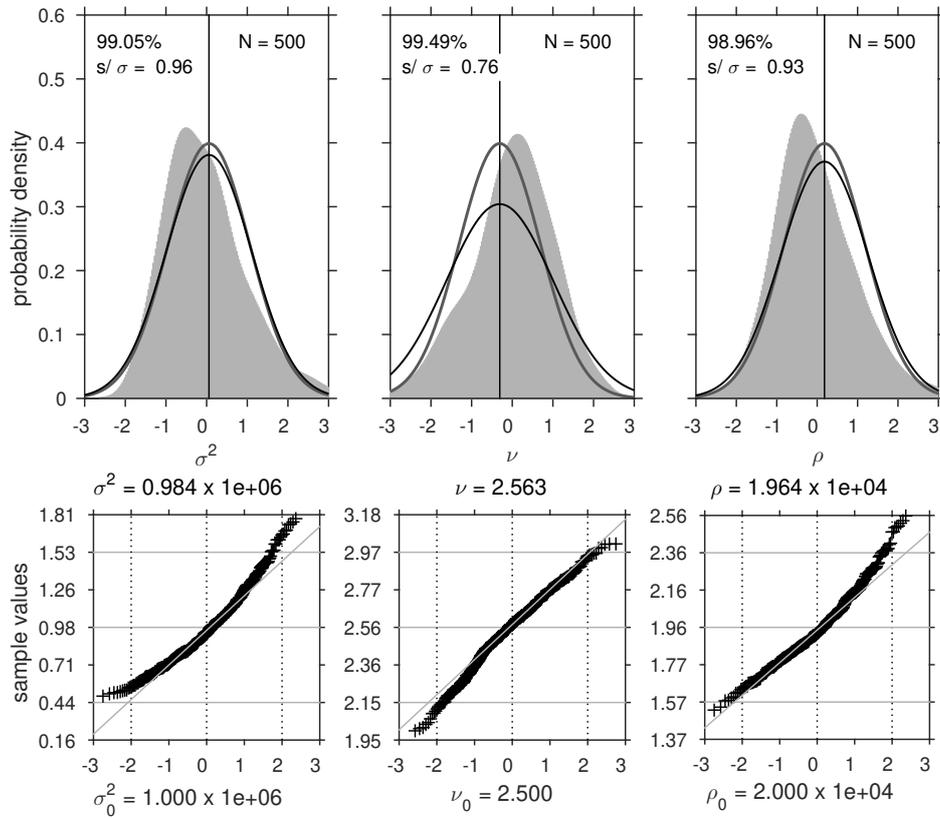


Figure 7. Recovery statistics of spatial-covariance embedding simulations carried out on a 64×64 grid, with spacings $\Delta x = \Delta y = 10$ km, and true values of $\sigma_0^2 = 1 \text{ km}^2$, $\nu_0 = 2.5$, $\rho_0 = 20$ km. The estimated parameters recovered over 500 simulations via the exactly blurred uncorrelated likelihood average to $\sigma^2 = 0.9841 \pm 0.2738 \text{ km}^2$, $\nu = 2.56 \pm 0.20$, $\rho = 19.641 \pm 1.986$ km. The thick gray line is derived from the covariance directly calculated from the ensemble of simulation and inversion outcomes. The thick black line is based on the covariance exactly calculated from eq. (53).

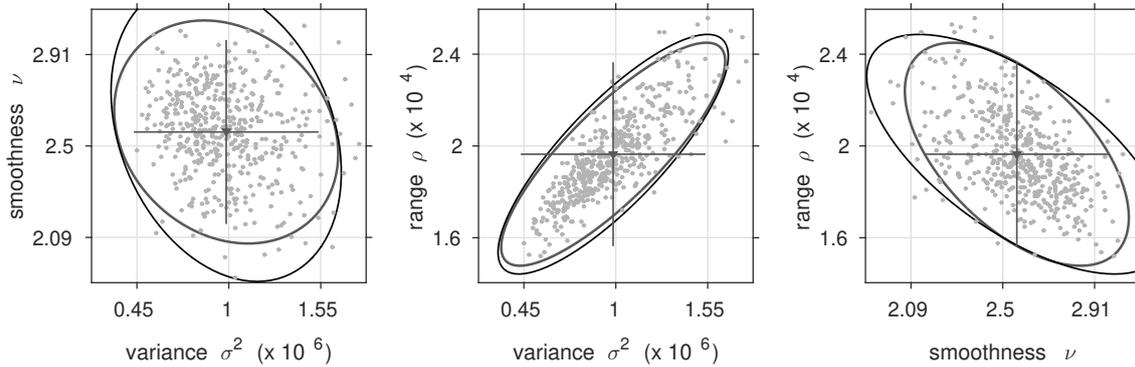


Figure 8. Very similar caption to the previous. Blabla. Same colors. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur condimentum venenatis magna. Ut vulputate urna vitae luctus sollicitudin. Etiam tempus nisl at mi imperdiet interdum. Pellentesque nec urna turpis. Etiam ultricies consectetur mauris hendrerit tristique. Suspendisse risus nisl, interdum sit amet euismod at, laoreet nec odio. Morbi eu sapien nec dui ultricies rhoncus vitae id mi. Phasellus pharetra, turpis vitae interdum fermentum, nisl est lobortis erat, vitae semper massa est quis ex. Nulla consequat.

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.

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 (43) may be. Failing the test could be due, for example, to the presence of patterns or preferred directions indicating that the data should rather be interpreted under anisotropic (e.g. geometric Goff & Jordan 1989b; Herzfeld & Overbeck 1999) extensions of the model. We save developing alternative hypotheses for future work.

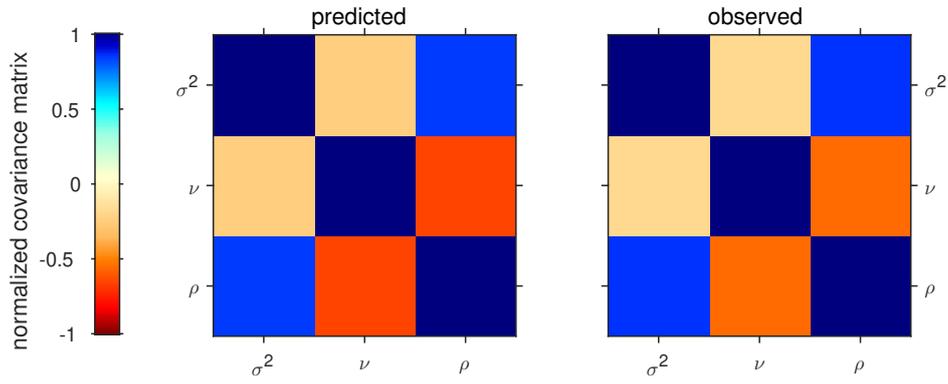


Figure 9. Covariance comparison. The numbers are as follows: observed $C_{\sigma^2\nu} = -0.1961$, $C_{\sigma^2\rho} = 0.8335$, $C_{\nu\rho} = -0.5678$, and predicted $C_{\sigma^2\nu} = -0.2837$, $C_{\sigma^2\rho} = 0.8167$, $C_{\nu\rho} = -0.6766$.

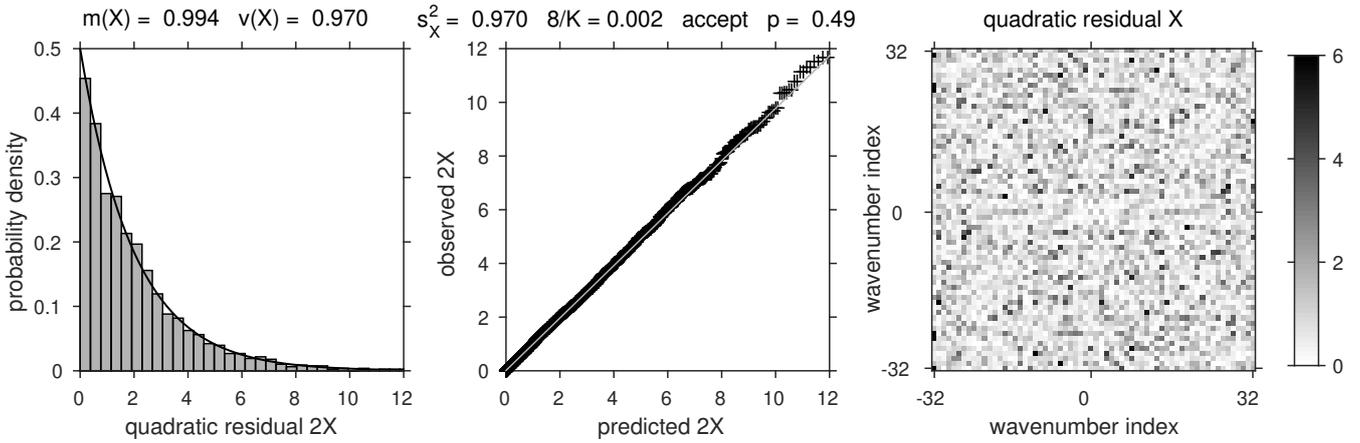


Figure 10. Residual statistics of simulations carried out on a 64×64 grid, with spacings $\Delta x = \Delta y = 10$ km, convolutional blurring, and true Matérn parameter values of $\sigma_0^2 = 1 \text{ km}^2$, $\nu_0 = 2.5$, $\rho_0 = 20$ km. Results pertaining to one of the simulations and its maximum-likelihood recovery. Distribution of the variable $X_\theta(\mathbf{k})$ of eq. (56), 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.

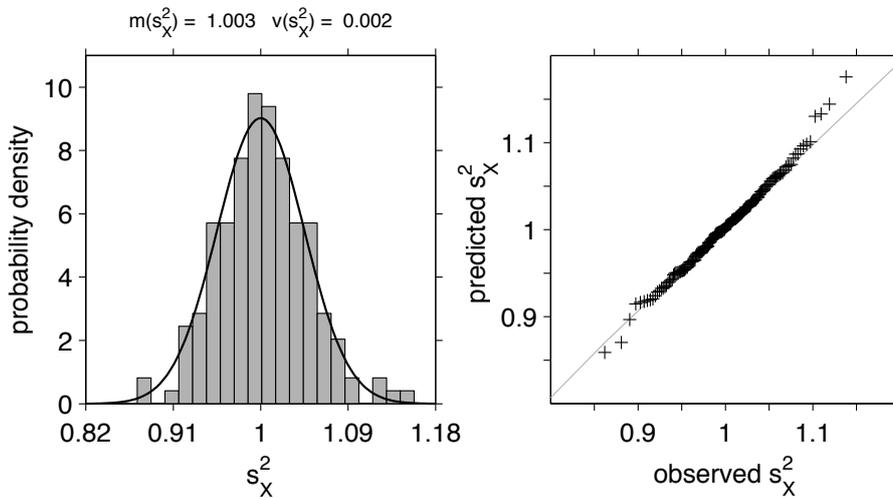


Figure 11. Report on the behavior of the test statistic s_X^2 of eq. (57) across an ensemble of 175 simulations and recovery experiments. Histogram and its prediction, quantile-quantile plot comparing observations to predictions, and spatial-domain map of one of the simulations. All of the theoretical behavior is validated; there are no surprises.

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

$$s_X^2 = \frac{1}{MN} \sum_{\mathbf{k}} [X_{\theta}(\mathbf{k}) - 1]^2. \quad (57)$$

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

$$\langle X_{\theta}^p \rangle = \Gamma(p + 1) = p!. \quad (58)$$

The case $p = 1$ discussed previously, in reducing eq. (40) to eq. (41), is easily verified. Evaluating eq. (58) for the case $p = 2$ then yields the expectation of eq. (57), our test statistic,

$$\langle s_X^2 \rangle = \frac{1}{MN} \sum_{\mathbf{k}} \langle X_{\theta}^2(\mathbf{k}) - 2X_{\theta}(\mathbf{k}) + 1 \rangle = 1. \quad (59)$$

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

$$MN \text{var}\{s_X^2\} = \text{var}\{X_{\theta}^2\} + 4 \text{var}\{X_{\theta}\} - 4 \text{cov}\{X_{\theta}^2, X_{\theta}\} \quad (60)$$

$$= \langle X_{\theta}^4 \rangle - \langle X_{\theta}^2 \rangle^2 + 4 \langle X_{\theta}^2 \rangle - 4 \langle X_{\theta} \rangle^2 - 4 \langle X_{\theta}^3 \rangle + 4 \langle X_{\theta}^2 \rangle \langle X_{\theta} \rangle = 8. \quad (61)$$

Hence we deduce that our chosen metric converges ‘in law’ to a variable distributed as:

$$s_X^2 \xrightarrow{\mathcal{L}} \mathcal{N}(1, 8/[MN]). \quad (62)$$

In other words, by computing eq. (57) 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 continues to show how the theory is borne out by simulation and experiment. The lower row of three panels reports on the ‘test of the test’, specifically, whether eq. (62) holds, across 175 simulations. Sample mean and variance of the test statistic are labeled above the first panel, which displays the histogram of s_X^2 over the simulations. The second panel shows the linearity of the quantile-quantile plots. Our conclusion is that using s_X^2 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.

5 PRACTICE

We are inching closer to the discussion of real-world data. Before we do so, we give one final example of residual analysis in the form of a diagnostic figure that we designed to be uncompromisingly informative yet cursorily consultable. We elect to show a *false* rejection, one of the 5 per cent such cases expected at the 95th per cent confidence level. The true Matérn parameter values were picked at random, as annotated, $\sigma_0^2 = 0.50 \text{ km}^2$, $\nu_0 = 2.70$, $\rho_0 = 60 \text{ km}$. The recovery, at $\sigma^2 = 0.49 \text{ km}^2$, $\nu = 2.65$, $\rho = 61 \text{ km}$, would have been deemed successful by conventional standards.

The leftmost two panels of Fig. 12 appear in a layout that is a combination of the top three panels of Fig. 11, including the histogram of $X_{\theta}(\mathbf{k}) = \bar{S}_{\theta}^{-1}(k) |H(\mathbf{k})|^2$ and its theoretical probability density (using the left ordinate) and quantile-quantile plot (using the right ordinate), and its wavenumber map. The rightmost panels show the blurred predicted spectral density $\bar{S}_{\theta}(\mathbf{k})$, and the observed periodogram $|H(\mathbf{k})|^2$. Contours of the predicted spectrum are drawn on the observed spectrum, showing a very good match.

Likelihood contours. Counterexample.

6 CONCLUSIONS

Rescue some of the high-falutin writing of the now mostly cut introduction.

What numbers capture the essence of a patch of planetary topography? Surprisingly few, as we show. Can we derive them, can we derive uncertainty bounds on them, can we simulate “new” realizations of patches that should look to the geologists exactly as if they were derived from similar processes? The answer to all of these questions is “yes”. Moreover, our approach holds in one, two, and three spatial dimensions, and it generalizes to multiple variables, e.g. when topography and gravity are being considered jointly (perhaps linked by flexural rigidity, erosion, or other surface and sub-surface modifying processes).

Equally important is our treatment of edge effects, both in simulation and in estimation. Our use of data tapering allows for the consideration of regions whose outlines are irregularly shaped, i.e. not rectangles or circles. Our procedure correctly determines the estimation variance of all of the parameters. Because even the maximum-likelihood “best” estimate may not be “good enough”, we developed tests on the residuals to ascertain whether the “model” itself fits or warrants rejection.

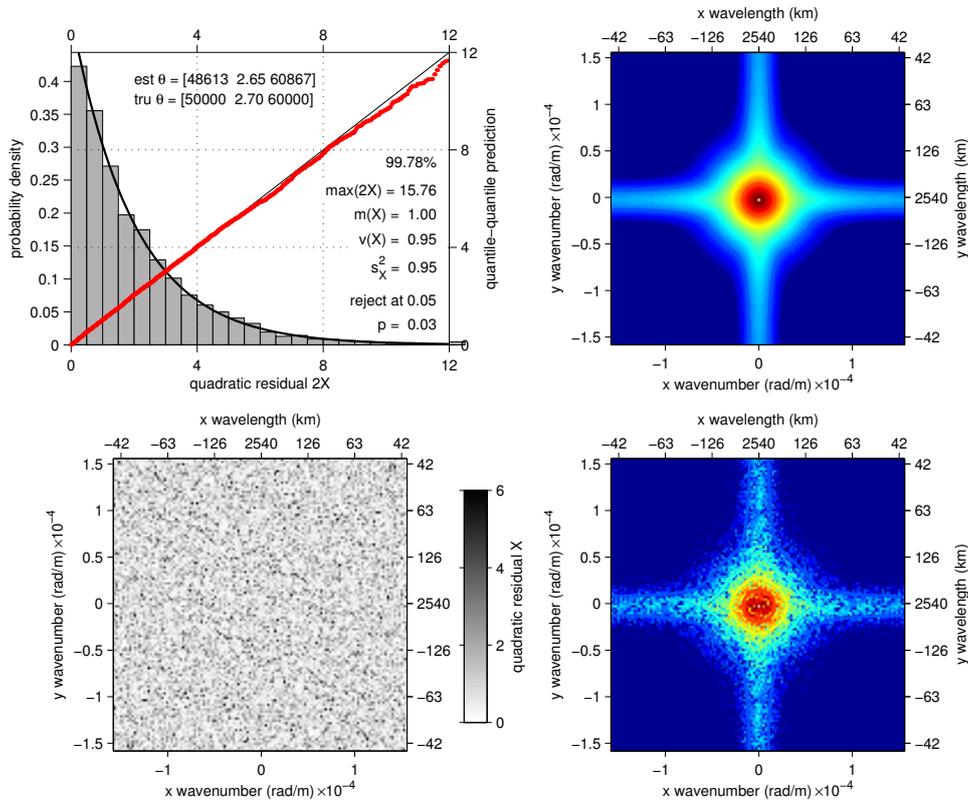


Figure 12. Example of a synthetic experiment where the maximum-likelihood estimate of simulated data ends up being close to the truth, and yet the model is rejected on the basis of hypothesis testing on the residual statistics. (*Top left:*) Histogram and its prediction, quantile-quantile plot, and (*Bottom right:*) spectral map of the residual variable $X_{\theta}(\mathbf{k}) = \bar{S}_{\theta}^{-1}(\mathbf{k})|H(\mathbf{k})|^2$, as in Fig. 11. (*Top right:*) Predicted blurred spectral density $\bar{S}_{\theta}(\mathbf{k})$. (*Bottom right:*) Observed periodogram $|H(\mathbf{k})|^2$, with contours of the prediction superposed.

7 ACKNOWLEDGEMENTS

This work was sponsored by the National Aeronautics and Space Administration under grant NNX11AQ45G. Our computer programs are available as SLEPIAN_Juliet at doi: NN. OLB 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 Studies, and Princeton University. We thank the Associate Editor, and the anonymous reviewers. Hi, Mom!

REFERENCES

- Abramowitz, M. & Stegun, I. A., 1965. *Handbook of Mathematical Functions*, Dover, 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.
- Bendat, J. S. & Piersol, A. G., 2000. *Random Data: Analysis and Measurement Procedures*, John Wiley, New York, 3rd edn.
- Cox, D. R. & Hinkley, D. V., 1974. *Theoretical Statistics*, Chapman and Hall, London, UK.
- Cramér, H., 1942. On harmonic analysis in certain functional spaces, *Arkiv Mat. Astr. Fysik*, **28B**, 1–7.
- Davison, A. C., 2003. *Statistical Models*, Cambridge Univ. Press, Cambridge, UK.
- Dietrich, C. R. & Newsam, G. N., 1997. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix, *SIAM J. Sci. Comput.*, **18**(4), 1088–1107, doi: 10.1137/S1064827592240555.
- 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.
- 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.

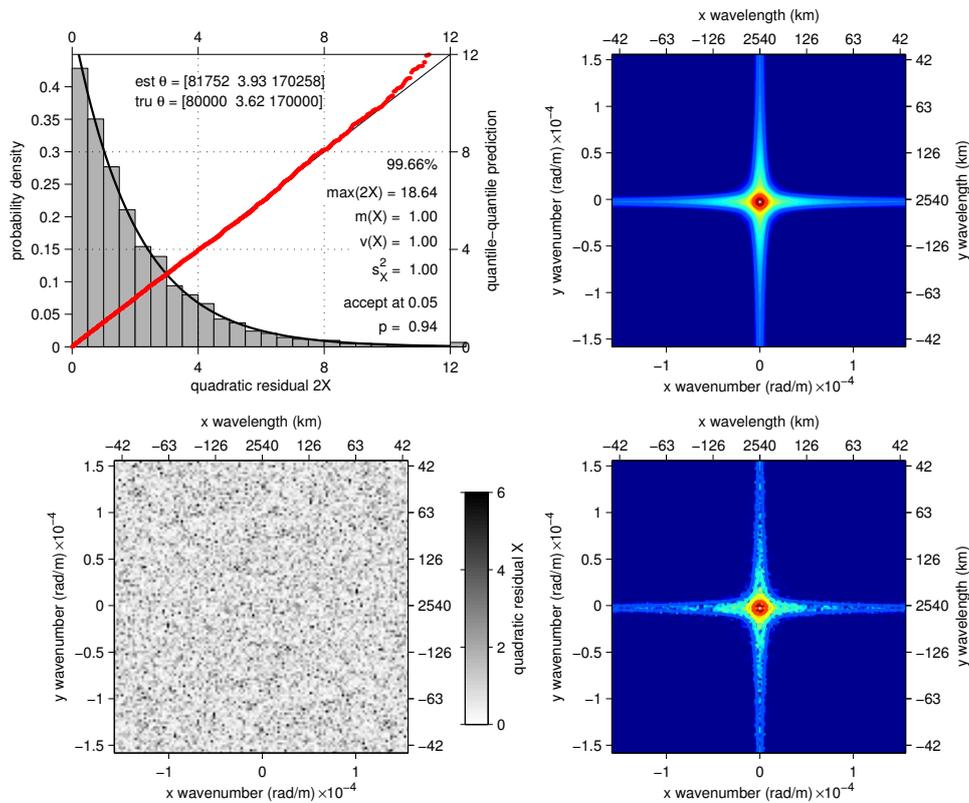


Figure 13. A synthetic, with the model accepted.

- 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.
- Handcock, M. S. & Stein, M. L., 1993. A Bayesian analysis of kriging, *Technom.*, **35**(4), 403–410.
- 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.
- 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.
- Mareschal, J.-C., 1989. Fractal reconstruction of sea-floor topography, *Pure Appl. Geophys.*, **131**(1–2), 197–210.
- 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.
- 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.
- 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.
- 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.
- 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.
- Stein, M. L., 1999. *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Series in Statistics, Springer, New York.
- 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.
- 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.
- Wood, A. T. A. & Chan, G., 1994. Simulation of stationary Gaussian processes in $[0, 1] d$, *J. Comput. Graph. Stat.*, pp. 409–432.

8 APPENDIX

For easy reference we list the expressions for the derivatives of the isotropic Matérn spectral density (12)–(13), in each of the three parameters $\theta \in \boldsymbol{\theta}$, via the parameter in which they enter the expressions (38)–(41) for the various derivatives of the likelihood (37), namely

$$m_{\theta}(k) = \mathcal{S}_{\boldsymbol{\theta}}^{-1}(k) \frac{\partial \mathcal{S}_{\boldsymbol{\theta}}(k)}{\partial \theta} = -\mathcal{S}_{\boldsymbol{\theta}}(k) \frac{\partial \mathcal{S}_{\boldsymbol{\theta}}^{-1}(k)}{\partial \theta} \quad (63)$$

Introducing the auxiliary variable

$$\mu = \left(\frac{4\nu}{\pi^2 \rho^2} \right) \left(\frac{4\nu}{\pi^2 \rho^2} + k^2 \right)^{-1}, \quad (64)$$

we can write the following three (which agree with eqs A25–A27 of Simons & Olhede 2013), which enter into the definition of the score (38) and the Fisher matrix (41),

$$m_{\sigma^2} = \frac{1}{\sigma^2}, \quad (65)$$

$$m_{\nu} = \frac{\nu+1}{\nu} - \frac{\nu+1}{\nu} \mu + \ln \mu, \quad (66)$$

$$m_{\rho} = -2 \frac{\nu}{\rho} + 2 \frac{\nu+1}{\rho} \mu. \quad (67)$$

The nonvanishing derivatives of eqs (65)–(67) necessary for the computation of the Hessian (40) are given by

$$\frac{\partial m_{\sigma^2}}{\partial \sigma^2} = -\frac{1}{\sigma^4}, \quad (68)$$

$$\frac{\partial m_{\nu}}{\partial \nu} = \frac{1}{\nu} \left(\frac{\nu+1}{\nu} - \mu + \frac{\nu+1}{\nu} \mu^2 \right), \quad (69)$$

$$\frac{\partial m_{\rho}}{\partial \rho} = 2 \frac{\nu}{\rho^2} \left(1 - 3 \frac{\nu+1}{\nu} \mu + 2 \frac{\nu+1}{\nu} \mu^2 \right), \quad (70)$$

$$\frac{\partial m_{\nu}}{\partial \rho} = \frac{2}{\rho} \left(-1 + \frac{2\nu+1}{\nu} \mu - \frac{\nu+1}{\nu} \mu^2 \right), \quad (71)$$