

Linking physics and spatial statistics: A new family of Boltzmann-Gibbs random fields

Denis Allard

INRAE, BioSP, 84914 Avignon, France
e-mail: denis.allard@inrae.fr

Dionisios T. Hristopoulos*

School of Electrical and Computer Engineering, Technical University of Crete, Chania
73100, Greece
e-mail: dchristopoulos@ece.tuc.gr

Thomas Opitz

INRAE, BioSP, 84914 Avignon, France
e-mail: thomas.opitz@inrae.fr

Abstract: We investigate a connection between spatial statistics and statistical physics to obtain new covariance functions with direct physical interpretation for spatial random fields. These covariance functions are based on the exponential Boltzmann-Gibbs representation and use an energy functional to represent interactions between the values of the random field at different points in space. This formulation results in closed-form generalized covariance functions, which display infinite variance in Euclidean spaces of dimension larger than one. We propose regularization schemes in real and reciprocal (spectral) space that lead to well-behaved covariance structures. The real-space regularization parameter allows a continuous interpolation between the Boltzmann-Gibbs covariance and the exponential covariance. We also propose discretized approximations on regular grids, and we show that they represent reparametrized versions of the well-known Besag and Leroux lattice models. We then discuss parameter estimation and spatial prediction for the regularized Boltzmann-Gibbs covariance model in two dimensions. We recommend using the pairwise difference likelihood that combines satisfactory estimation performance and good scalability with many observation points. The predictive performance of the regularized covariance function is assessed by means of cross-validation statistics. Irregularly-spaced samples from the Walker Lake dataset are used, and spatial prediction is conducted by means of ordinary kriging. The regularized Boltzmann-Gibbs covariance yields improved predictive performance compared to the exponential covariance model.

MSC2020 subject classifications: Primary 60K35, 62H11, 62M30; secondary 62P12, 62P35.

Keywords and phrases: Besag process, estimation, Leroux process, prediction, regularization, spatial covariance.

Received February 2021.

*Corresponding author. Tel: +030-28210-37688, Fax: +030-28210-37853.

Contents

1	Introduction	4086
2	Boltzmann-Gibbs random fields	4088
2.1	Energy representation	4088
2.2	Spectral representation	4090
2.3	Covariance representation	4091
2.4	Some special cases	4092
3	Regularization of the covariance function	4093
3.1	Real-space regularization	4094
3.1.1	Theoretical result	4094
3.1.2	Variograms of regularized BG random fields	4095
3.2	Frequency-space regularization	4097
3.3	Matching the regularization schemes	4100
4	Statistical models for lattice data	4101
4.1	Energy terms for lattice models	4102
4.2	Links to common lattice models	4103
5	Parameter estimation	4104
6	Case study: Walker Lake data	4105
6.1	Description of the data set	4105
6.2	Estimation of the variogram	4106
6.2.1	Estimation with full dataset	4106
6.2.2	Estimation using smaller training sets	4107
6.3	Prediction and cross-validation	4110
7	Conclusions	4112
	References	4113

1. Introduction

There are strong links between statistical physics and spatial statistics. These date back to the work of Besag [6, 7] on lattice autoregression models, which was inspired by the Ising model developed for magnetic systems by the physicist Ernst Ising [32], and to the seminal Monte Carlo algorithm of Nicholas Metropolis [38], which was initially developed for the simulation of liquids and today occupies a central position in Bayesian statistics and simulation-based statistical inference [42].

In statistical physics, the spatial dependence and variation of field values is usually expressed in terms of Boltzmann-Gibbs probability density functions (pdf), which have the general form $f = \exp(-\mathcal{H})/Z$, where \mathcal{H} is an energy functional, and Z is a normalization constant known as the partition function. For example, this formulation is used in statistical field theory [34]. The partition can be explicitly calculated in finite-dimensional cases if the model is Gaussian, i.e., if the energy functional is of quadratic form, but it is usually intractable in more general cases. Nonetheless, various techniques allow the approximate calculation of Z [26].

The main ideas of statistical field theory have been used to develop Boltzmann-Gibbs random fields with a new class of covariance structure [24, 28, 25], including cross-covariance models for multivariate data [29]. Furthermore, combining the equilibrium (i.e., purely spatial) Boltzmann-Gibbs fields with linear response theory, a new space-time covariance function was obtained [30].

The Boltzmann-Gibbs framework was also used to develop the so-called stochastic local interaction (SLI) model. This model defines a Gaussian Markov random field over the observed locations, which may be irregularly spaced, by using an appropriately specified neighborhood for each location. Then, fast numerical estimation and prediction are possible even with large numbers of locations, and without resorting to the framework of finite-element-based approximations to stochastic partial differential equations (SPDEs) [37]. The SLI model has been recently extended to the spatiotemporal setting [27]. The main advantage of the Boltzmann-Gibbs representation in this case is that it provides *explicit and sparse* expressions for the precision matrix (i.e., the inverse covariance matrix), which partake a significant computational advantage with respect to the classical, covariance-based approaches (e.g., kriging, and Gaussian process regression).

In this study, we investigate theoretical properties and statistical techniques for a form of Gaussian Boltzmann-Gibbs models whose energy functional \mathcal{H} is related to the spatial variation of the realizations of the field $x(\mathbf{s})$ and of its gradient function $\nabla x(\mathbf{s})$. High energies (and therefore low probabilities of realization) are associated with large values of the squares of the field and its gradient. Two parameters are used to scale the magnitude of the contributions from these two terms to the overall energy. This approach offers a natural mechanism for assigning lower probability of realization to functions $x(\mathbf{s})$ that show strong global or local variability in space, such that spatial coherence of values $x(\mathbf{s})$ across different locations \mathbf{s} is established and can be controlled parametrically.

The content of the remainder of the paper is structured as follows. In Section 2, we define the Boltzmann-Gibbs energy with respect to an appropriate function space as starting point and then derive distributional properties of the resulting random fields. Specifically, we characterize the spatial variability through both its spectral representation and the corresponding covariance functions. This allows highlighting links to existing random field models that have been explored for statistical implementations. In Euclidean space of dimension larger than one, our formulation yields generalized covariances with infinite variance. Therefore, in Section 3 we develop and compare two regularization schemes that define approximating models with finite variance. One of the regularization schemes operates in the spectral domain and the other directly on the covariance function. For both approaches, we analytically and numerically investigate the behavior of the variogram function in different limits. Section 4 highlights links to the Besag and Leroux models on regular lattices, which can be viewed as another regularization approach. Section 5 discusses parameter estimation of the regularized covariance function using maximum likelihood techniques. A simulation study illustrates the performance of the estimators for various data configurations based on subsamples of the Walker Lake dataset; we advocate

the pairwise difference likelihood as a method that provides accurate estimations and good scalability with many observation locations. Section 6 presents a study to assess the prediction performance of the new covariance models using the Walker Lake dataset. Finally, a discussion of the results and our conclusions are presented in Section 7.

2. Boltzmann-Gibbs random fields

We consider spatial models defined over a domain \mathcal{D} which we choose as a compact subset of \mathbb{R}^d , where $\mathbf{s} \in \mathbb{R}^d$ denotes the location vector. The boundary $\partial\mathcal{D}$ of \mathcal{D} is supposed to be a piecewise smooth (i.e., infinitely differentiable) manifold of dimension $d - 1$. More precisely, $\{X(\mathbf{s}, \omega) : \mathbf{s} \in \mathbb{R}^d, \omega \in \Omega\}$ is a scalar, real-valued *spatial random field* defined on a probability space (Ω, \mathcal{F}, P) , where Ω is the sample space, \mathcal{F} is a σ -field of subspaces of Ω , and P is a probability measure [49]. The realizations of the random field $X(\mathbf{s}; \omega)$ for $\omega \in \Omega$ are deterministic functions denoted by $x(\mathbf{s})$. The expectation over all states will be denoted by $\mathbb{E}[\cdot]$, and $m_x(\mathbf{s}) = \mathbb{E}[X(\mathbf{s}; \omega)]$ if the expectation is well defined, i.e., if $\int_{\Omega} |x(\mathbf{s}; \omega)| P(d\omega) < \infty$. Without loss of generality, we will consider below $m_x(\mathbf{s}) = 0$ unless otherwise specified.

2.1. Energy representation

In this work, we consider realizations $x(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^d$, possessing regularity properties that give mathematical sense to operations such as differentiation (up to the second order), integration on compact subsets of \mathbb{R}^d and Fourier transforms, i.e., we ensure that such operations are well-defined. In the usual mean-square sense, these regularity conditions are too restrictive for practical use. We therefore consider them in a *weak sense*. In recent work, [11] defined Generalized Random Fields as tempered distributions from Schwartz's theory of distributions [45]. In this framework, the operations of differentiation, integration and Fourier Transform are well defined and are considered in a distributional sense. Here, we will suppose that $x(\mathbf{s})$ belongs to the Sobolev space $\mathcal{W}^{2,2}(\mathcal{D})$, which is the weak sense restriction to \mathcal{D} of the generalized functions (i.e. tempered distributions) on \mathbb{R}^d admitting square integrable second-order distributional derivatives. In $\mathcal{W}^{2,2}(\mathcal{D})$, point values such as $x(\mathbf{s})$ or the Laplacian of the function $x(\mathbf{s})$,

$$\nabla^2 x(\mathbf{s}) = \sum_{i=1}^d \frac{\partial^2 x(\mathbf{s})}{\partial s_i^2},$$

are expressed in terms of integrable generalized functions, such as Dirac delta functions. For simplicity of exposition, we abuse notation and conveniently write $x(\mathbf{s})$ and $\nabla^2 x(\mathbf{s})$ to refer to these tempered distributions.

The focus of our work is on random fields characterized by the Boltzmann-Gibbs probability density function (pdf) $f = \exp(-\mathcal{H})/Z$ over $\mathcal{W}^{2,2}(\mathcal{D})$ given

by

$$\mathcal{H}[x_{\mathcal{D}}; \boldsymbol{\theta}] = \frac{1}{2} \int_{\mathbb{R}^d} \left\{ \alpha_0 x^2(\mathbf{s}) + \alpha_1 [\nabla x(\mathbf{s})]^2 \right\} \mathbb{I}(\mathbf{s} \in \mathcal{D}) d\mathbf{s}, \tag{2.1}$$

where $x_{\mathcal{D}} = \{x(\mathbf{s}), \mathbf{s} \in \mathcal{D}\}$, $\boldsymbol{\theta} = (\alpha_0, \alpha_1)^\top$ with $\alpha_0, \alpha_1 \in \mathbb{R}_+^*$ (where \mathbb{R}_+^* is the set of positive real numbers), $\nabla = \left(\frac{\partial}{\partial s_1}, \dots, \frac{\partial}{\partial s_d}\right)$ is the gradient operator and $\mathbb{I}(A)$ is the indicator function with $\mathbb{I}(A) = 1$ if the condition A is satisfied and $\mathbb{I}(A) = 0$ otherwise.

In the pdf, the denominator Z is the normalizing constant, frequently referred to as the *partition function*. The functions $x(\mathbf{s})$ belong to the space $\mathcal{W}^{2,2}(\mathcal{D})$, and they denote realizations (states) of the random field $X(\mathbf{s}; \omega)$. We shall refer to such random fields as *Boltzmann-Gibbs (BG) Random Fields*.

From a rigorous, mathematically oriented perspective, the representation (2.1) should be set in the framework of Gaussian measures on Hilbert spaces [13]. The probability density would then be deduced from the respective characteristic function, involving its trace operator on $\mathcal{W}^{2,2}(\mathcal{D})$. This rigorous approach would come at the cost of highly technical functional analysis arguments. Since the focus of this work is on the statistical use of BG random fields, we opt for a more direct, physics-oriented characterization. With this perspective in mind, the main properties of BG random fields $X(\mathbf{s}; \omega)$ can be stated as follows:

- (A.1) The partition function $Z = \int_{\mathcal{W}^{2,2}(\mathcal{D})} \exp(-\mathcal{H}(x_{\mathcal{D}})) dx_{\mathcal{D}}$ involves the functional integral over all states $x(\mathbf{s})$ [18, 34, 3, 26]. This represents an integral over an infinite-dimensional space, the rigorous definition of which is rather involved [3]. However, for our purposes, the mathematical details arising in the calculation of Z are circumstantial since the statistical properties of the random field follow directly from the structure of the energy functional without making explicit use of Z .
- (A.2) The random field is *Gaussian*. This property derives from the fact that the energy is a quadratic functional of $x(\mathbf{s})$.
- (A.3) The *expectation is equal to zero*. This property follows from the fact that for Gaussian random fields the expectation coincides with the mode of the pdf. Since $\alpha_0, \alpha_1 > 0$, it holds that $\mathcal{H} \geq 0$, and the mode is obtained for $\mathcal{H} = 0$. However, we get $\mathcal{H} = 0$ if and only if $x(\mathbf{s})$ is the zero element in $\mathcal{W}^{2,2}(\mathcal{D})$, which corresponds to the generalized version in $\mathcal{W}^{2,2}(\mathcal{D})$ of the null function over \mathbb{R}^d .
- (A.4) The random field is *stationary*. This follows from the fact that (i) the expectation is zero, and (ii) the parameters α_0 and α_1 , which determine the generalized covariance function (see Section 2.2), are independent of the spatial locations. In particular, the energy functional admits a spectral representation, cf. Eq. (2.6), which involves a translation-invariant spectral density determined by α_0 and α_1 . Since the field is Gaussian, the translation invariance of the mean and the covariance function guarantees both wide-sense and strict-sense stationarity.
- (A.5) The random field is *statistically isotropic*. This property reflects the fact that the energy functional does not have a directional preference. The

isotropy is broken if, e.g., $(\nabla x(\mathbf{s}))^2$ is replaced by $(\mathbf{b} \cdot \nabla x(\mathbf{s}))^2$, where $\mathbf{b} \in \mathbb{R}^d$ and $\mathbf{b} \cdot \nabla x(\mathbf{s})$ denotes the dot (inner) product of the two vectors.

In more general cases where we consider random fields whose mean function $m_x(\mathbf{s})$ is not zero everywhere, we assume that it also belongs to $\mathcal{W}^{2,2}(\mathcal{D})$. If the mean function is not constant, $x(\mathbf{s})$ includes a non-stationarity in the mean. Stationarity and isotropy then hold for $x(\mathbf{s}) - m_x(\mathbf{s})$ in (2.1). From now on, $m_x(\mathbf{s})$ is assumed to be equal to zero everywhere.

In the following, the integral of the square gradient in (2.1) is expressed in terms of the Laplacian of $x(\mathbf{s})$. Let dS be the unit surface element on $\partial\mathcal{D}$, and $\mathbf{n}(\mathbf{s})$ the unit outward-pointing vector perpendicular to $\partial\mathcal{D}$ for $\mathbf{s} \in \partial\mathcal{D}$. Then, *Green's first identity*, which extends integration by parts to dimensions $d > 1$, leads to the following representation:

$$\int_{\mathcal{D}} (\nabla x(\mathbf{s}))^2 d\mathbf{s} = - \int_{\mathcal{D}} x(\mathbf{s}) \nabla^2 x(\mathbf{s}) d\mathbf{s} + \int_{\partial\mathcal{D}} x(\mathbf{s}) [\nabla \mathbf{n}(\mathbf{s}) \cdot \nabla x(\mathbf{s})] dS. \quad (2.2)$$

The standard assumption in statistical field theory is that the boundary term, i.e., the last integral on the right hand side of Eq. (2.2), can be neglected [19]. Indeed, if the domain \mathcal{D} is a hypercube $[0, L]^d$, the ‘‘bulk’’ term, i.e., the first integral on the right hand side of Eq. (2.2), extends over a domain that scales as L^d , while the boundary term involves an integral over a domain that scales as L^{d-1} . Hence, if $L \gg 1$, the contribution of the boundary term becomes negligible in comparison to the bulk.

In light of the above, we propose to work with the energy functional given as

$$\mathcal{H}[x_{\mathcal{D}}; \boldsymbol{\theta}] = \frac{1}{2} \int_{\mathbb{R}^d} [\alpha_0 x^2(\mathbf{s}) - \alpha_1 x(\mathbf{s}) \nabla^2 x(\mathbf{s})] \mathbb{I}(\mathbf{s} \in \mathcal{D}) d\mathbf{s}. \quad (2.3)$$

In the Boltzmann-Gibbs approach, the regularity properties of the realizations (i.e., of the sample paths) are not derived from the covariance function, as is commonly done with mathematical representations of Gaussian random fields based on the covariance [2]. In contrast, our approach consists of defining the class of functions (i.e., sample paths) admissible in the energy functional (2.1), and the properties of the joint pdf and the covariance function then follow.

2.2. Spectral representation

In this section we derive a spectral representation for the energy functional given by Eq. (2.1) in terms of the *Fourier transform* (FT). The Fourier representation of $x(\mathbf{s})\mathbb{I}(\mathbf{s} \in \mathcal{D})$ involves a series of discrete frequency vectors \mathbf{k}_i . The series representation converges to the Fourier integral as $\mathcal{D} \rightarrow \mathbb{R}^d$. However, the functions $x(\mathbf{s})$ in the energy functional Eq. (2.1) are not absolutely integrable over \mathbb{R}^d . Hence, their Fourier transform does not exist in the ordinary sense. Nonetheless, a generalized Fourier transform can be defined for tempered distributions by means of suitably defined test functions [11, 17, 41]. For the sake

of brevity, we shall directly use the main results of this theory regarding Fourier transforms. Interested readers will find all the theoretical justifications in the references above. In the following, $\tilde{x}(\mathbf{k})$ refers to the generalized Fourier transform. The BG random field states can then be expressed using the weak-sense Cramér representation of $x(\mathbf{s})$, given in terms of the following integral

$$x(\mathbf{s}) = \text{IFT}[\tilde{x}(\mathbf{k})] = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{j\mathbf{k}\cdot\mathbf{s}} \tilde{x}(\mathbf{k}) \, d\mathbf{k}, \tag{2.4}$$

where j is the imaginary unit, $\mathbf{k} \in \mathbb{R}^d$ denotes the *frequency vectors* in reciprocal (Fourier) space, and IFT denotes the inverse Fourier transform.

The energy functional can be expressed in the spectral representation. For this, we take into account the orthonormality of the Fourier basis, i.e.,

$$\int_{\mathbb{R}^d} e^{j(\mathbf{k}+\mathbf{k}')\cdot\mathbf{s}} \, d\mathbf{s} = (2\pi)^d \delta(\mathbf{k} + \mathbf{k}'),$$

and the algebraic representation of the second-order partial derivatives in Fourier space, i.e., $\text{FT}[\partial^2 x(\mathbf{s})/\partial s_i^2] = -k_i^2 \tilde{x}(\mathbf{k})$, which leads to $\text{FT}[\nabla^2 x(\mathbf{s})] = -\|\mathbf{k}\|^2 \tilde{x}(\mathbf{k})$. We thus obtain the following expression of the energy functional in (2.1) which corresponds to Parseval’s identity (i.e., the equality of the energy content in the direct and reciprocal spaces):

$$\mathcal{H}[\tilde{x}_{\mathcal{D}}; \boldsymbol{\theta}] = \frac{1}{2(2\pi)^d} \int_{\mathbb{R}^d} \tilde{x}^*(\mathbf{k}) [\alpha_0 + \alpha_1 \|\mathbf{k}\|^2] \tilde{x}(\mathbf{k}) \, d\mathbf{k}, \tag{2.5}$$

where $\tilde{x}_{\mathcal{D}} = \{\tilde{x}(\mathbf{k}), \mathbf{k} \in \mathbb{R}^d\}$ and z^* represents complex conjugation of z . Based on this expression, the *BG spectral density* is given by

$$\tilde{C}(\mathbf{k}) = \frac{1}{\alpha_0 + \alpha_1 \|\mathbf{k}\|^2} = \frac{1}{\alpha_0 \left(1 + \frac{\alpha_1}{\alpha_0} \|\mathbf{k}\|^2\right)}. \tag{2.6}$$

Hence $a = \sqrt{\alpha_1/\alpha_0}$ can be interpreted as the (*length*) *BG scale parameter*, whereas $1/(\alpha_0 a^d)$ corresponds to the *BG variance parameter*, which we will denote by σ^2 from now on.

2.3. Covariance representation

The covariance function is obtained through the following inverse Fourier transform (where $\mathbf{h} \in \mathbb{R}^d$ is the *vector-valued lag* in real space):

$$C_d(\mathbf{h}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{j\mathbf{k}\cdot\mathbf{h}} \tilde{C}(\mathbf{k}) \, d\mathbf{k}. \tag{2.7}$$

Since $\tilde{C}(\mathbf{k})$ is a radial function (i.e., its dependence on \mathbf{k} is only through $k = \|\mathbf{k}\|$), the covariance function is also radial, i.e., it depends on \mathbf{h} only through the Euclidean distance $r = \|\mathbf{h}\| \geq 0$. For the sake of conciseness, we will use the same

notation C for the vector covariance function and its radial representation. The *radial spectral representation* is given by the following *Hankel transform* [49, 26]:

$$C_d(r) = \frac{1}{(2\pi)^{d/2} r^\nu} \int_0^\infty k^{\nu+1} J_\nu(kr) \tilde{C}(k) dk, \quad (2.8)$$

where $J_\nu(\cdot)$ is the Bessel function of the first kind and order $\nu = d/2 - 1$. Using the BG spectral density (2.6) and provided that $\alpha_1 > 0$, the spectral representation becomes

$$C_d(r) = \frac{1}{(2\pi)^{d/2} \alpha_1 r^\nu} \int_0^\infty \frac{k^{\nu+1} J_\nu(kr)}{\frac{\alpha_0}{\alpha_1} + k^2} dk.$$

The above integral can be evaluated using the Hankel-Nicholson integration formula [1, p. 488], [20, p. 306, Eq. 6.565.4], which leads to the following expression for $-1 < \nu < 3/2$ which is satisfied for $d \in \{1, 2, 3, 4\}$:

$$C_d(r) = \frac{1}{(2\pi)^{d/2} \alpha_1} \left(\frac{1}{ar} \right)^\nu K_\nu(r/a). \quad (2.9)$$

In Eq. (2.9), $K_\nu(\cdot)$ is the *modified Bessel function of the second kind* of order ν . In particular, it follows that the generalized BG covariance is given by

$$C_d(r) = \begin{cases} \frac{1}{2a\alpha_0} e^{-r/a} & = \frac{\sigma^2}{S_1} e^{-r/a}, & d = 1, \\ \frac{1}{2\pi\alpha_1} K_0(r/a) & = \frac{\sigma^2}{S_2} K_0(r/a), & d = 2, \\ \frac{1}{4\pi\alpha_1 r} e^{-r/a} & = \frac{\sigma^2}{S_3} \frac{\exp(-r/a)}{r/a}, & d = 3, \\ \frac{1}{4\pi^2\alpha_1 ar} K_1(r/a) & = \frac{\sigma^2}{2S_4} \frac{K_1(r/a)}{r/a} & d = 4. \end{cases} \quad (2.10)$$

Above, $S_d = 2\pi^{d/2}/\Gamma(d/2)$ denotes the surface area of the $d - 1$ -dimensional unit sphere (i.e., the “area” of the boundary of the d -dimensional unit ball), with $S_1 = 2$, $S_2 = 2\pi$, $S_3 = 4\pi$ and $S_4 = 2\pi^2$. To see how the expressions for $d = 1$ and $d = 3$ follow from Eq. (2.9), we need to take into account two aspects: (i) $K_{-1/2}(x) = K_{1/2}(x)$, based on the symmetry of the modified Bessel function of the second kind with respect to ν [26, p. 336] [4, p. 686], and (ii) $K_{1/2}(x) = \sqrt{\frac{\pi}{2x}} e^{-x}$; the latter follows from the integral representation of $K_\nu(x)$ for $\nu = 1/2$ [4, Eq. (14.131)].

In Eq. (2.10), we see that $C_2(\cdot)$ to $C_4(\cdot)$ exhibit a singularity at $r = 0$. More generally, for $d = 2, 3, 4$ the function $C_d(\mathbf{h})$ is a *generalized covariance* [28, 11]. This is a consequence of the fact that both the BG field’s derivatives and its Fourier transform are defined in the generalized sense.

2.4. Some special cases

We now elaborate on two special cases of the BG random field.

Whittle-Matérn (W-M) random fields: The spectral density of Whittle-Matérn (W-M) random fields is given by $\tilde{C}(\mathbf{k}) = A_{\nu,d}/(1 + \|\mathbf{k}\|^2 a^2)^{\nu+d/2}$, where $A_{\nu,d}$ is a coefficient that depends on the dimension of space and the smoothness index $\nu > 0$ of the W-M field. A comparison between the above and Eq. (2.6) shows that the latter describes a W-M covariance with $\nu = 1/2$ in $d = 1$. In $d = 2$, the BG spectral density is equivalent to the W-M spectral density for $\nu = 0$. As shown in [11], the random field corresponding to $\nu = 0$ is a Generalized Random Field, which is well defined in the framework of tempered random distributions. The associated covariance is a covariance distribution, referred to as the generalized Matérn covariance. In [37, 11], it was shown that this Generalized Random Field arises as a weak solution of the stochastic partial differential equation (SPDE)

$$(\kappa^2 - \Delta)^{1/2}x(\mathbf{s}) = \varepsilon(\mathbf{s}), \quad \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^2, \quad (2.11)$$

where $\kappa^2 > 0$, $\varepsilon(\mathbf{s})$ is Gaussian white noise, and the SPDE is fractional due to the non-integer exponent $1/2$. Using appropriate boundary conditions on the PDE of Neumann or Dirichlet type, [37] defined a stochastic weak solution and obtained well-defined and accurate approximations through classical Gaussian random fields. Then, the link to the parametrization of $C_2(r)$ in (2.10) is given by $a = 1/\kappa$.

Scale-free case: For $\alpha_0 = 0$, the BG spectral density becomes $\tilde{C}(\mathbf{k}) = (\alpha_1 \|\mathbf{k}\|^2)^{-1}$. This is the Fourier representation of the generalized covariance function $K(r) = -A_d/r^{d-2}$ for $d \geq 3$ and $K(r) = -A_2 \log r$ for $d = 2$. In this case, the covariance does not have a characteristic scale a . The generalized covariance function is undefined at $r = 0$, but it is still positive definite in a distributional sense. Furthermore, it represents the Green's function of the Laplace equation [33]. In addition, the logarithmic function obtained in $d = 2$ corresponds to the *De Wijs* geostatistical model [39, 12]. Again, this random field can be interpreted as a stochastic weak solution to the SPDE in (2.11) when $\kappa^2 = 0$; see [37, 11] for details.

3. Regularization of the covariance function

The covariances obtained for $d = 2$ to $d = 4$, i.e., $C_2(\cdot)$, $C_3(\cdot)$ and $C_4(\cdot)$, exhibit a singularity at zero which hinders their practical use. In this section, we use regularization to build easily tractable finite-variance approximations, which can be used as standard stationary covariance functions for prediction and simulation. The term “regularization” is here used for procedures that lead to stable solutions of ill-posed problems [23]. In the case of divergent Feynman integrals obtained in statistical field theory, various regularization techniques can be used to make the integrals finite, including momentum-cutoff (spectral-cutoff), the Pauli-Villars method, as well as lattice-based and dimensional regularization methods [35]. Herein, we consider two options: real-space regularization using a spatial cutoff at very small distances, and regularization with a spectral cutoff at large frequencies.

3.1. Real-space regularization

In this section we introduce a regularization method which is based on the use of a small *cutoff distance* ε .

3.1.1. Theoretical result

We first state the general theoretical result in the following proposition and then provide its proof. The set of positive integers is denoted \mathbb{N}^* .

Proposition 3.1. *Let $\varepsilon > 0$. The function $C_{a,\varepsilon}(\mathbf{h}) = C_a(r + \varepsilon)$, where $r = \|\mathbf{h}\|$, $\mathbf{h} \in \mathbb{R}^\ell$ and $C_a(\cdot)$ is any of the radial forms defined on the right-hand side of (2.10), is positive definite in \mathbb{R}^ℓ for any $\ell \in \mathbb{N}^*$.*

For the proof, the following well-known proposition recalls some properties characterizing radial functions that are strictly positive definite in \mathbb{R}^ℓ for any $\ell \in \mathbb{N}^*$. Proofs can be found in [47, 15].

Proposition 3.2 (Completely monotone functions). *For a radial function $\varphi(r) : [0, \infty] \rightarrow \mathbb{R}$ the following three properties are equivalent:*

1. $\varphi(\|\mathbf{h}\|^2)$ is strictly positive definite on \mathbb{R}^ℓ for every $\ell \in \mathbb{N}^*$;
2. $\varphi(r)$ is completely monotone on $[0, \infty)$ and non constant;
3. there exists a non negative Borel measure μ on $[0, \infty)$ that does not concentrate on 0 and satisfies $\varphi(r) = \int_0^\infty e^{-tr} \mu(dt)$.

The equivalence of the properties (2) and (3) is known as the Bernstein theorem [5], while the equivalence of (1) and (2) was shown by Schoenberg [44]. The equivalence of (1) and (3) follows from the other two equivalences. We also need the following Lemma, which is an interesting result on its own.

Lemma 3.1. *Let $\varphi(z)$ be a completely monotone (CM) function on $(0, \infty)$ or on $[0, \infty)$. Then the function $\varphi_\varepsilon(z) := \varphi(z + \varepsilon)$ with $\varepsilon > 0$ is CM on $[0, \infty)$.*

Proof. As stated in Proposition 3.2, the function $\varphi(z)$ is a CM function on $(0, \infty)$ (respectively on $[0, \infty)$), if and only if there exists a positive measure μ on $[0, \infty)$ such that for all $z > 0$ (respectively $z \geq 0$),

$$\varphi(z) = \int_0^\infty e^{-zt} \mu(dt).$$

Let us now consider the positive measure $g(dt) = e^{-\varepsilon t} \mu(dt)$ with $\varepsilon > 0$. Then,

$$\int_0^\infty e^{-zt} g(dt) = \int_0^\infty e^{-(z+\varepsilon)t} \mu(dt) = \varphi_\varepsilon(z),$$

which shows that φ_ε is a completely monotone function. Moreover, since $\varphi_\varepsilon(0) = \varphi(\varepsilon) < \infty$, it is defined on $[0, \infty]$. \square

Equipped with these preliminary results we are now able to prove the main result in Proposition 3.1.

Proof of Proposition 3.1. The negative exponential is known to be positive definite on \mathbb{R}^ℓ for any $\ell \in \mathbb{N}^*$. Following [14, Proposition 4], the function $z^{\nu/2}K_\nu(\sqrt{z})$ is CM on $(0, \infty)$. By application of Lemma 3.1, the function $(z+\varepsilon)^{\nu/2}K_\nu(\sqrt{z+\varepsilon})$ with $\varepsilon > 0$ is thus CM on $[0, \infty)$ since it is finite at $z = 0$. Hence, the function $(\|\mathbf{h}\| + \varepsilon)^\nu K_\nu(\|\mathbf{h}\| + \varepsilon)$ is positive definite on \mathbb{R}^ℓ , for any $\ell \in \mathbb{N}^*$. By continuity, $\lim_{\nu \rightarrow 0} (\|\mathbf{h}\| + \varepsilon)^\nu K_\nu(\|\mathbf{h}\| + \varepsilon) = K_0(\|\mathbf{h}\| + \varepsilon)$ is thus also positive definite on \mathbb{R}^ℓ , which proves the Proposition for $d = 2$. For $d \geq 3$, we first use the fact that $z^{-\nu}$ is CM on $(0, \infty)$ when $\nu > 0$. Thus, the product $z^{-\nu} z^{\nu/2} K_\nu(\sqrt{z}) = z^{-\nu/2} K_\nu(\sqrt{z})$, is also CM on $(0, \infty)$ since the product of two CM functions is CM. Then, the same reasoning as above applies: by application of Lemma 3.1, the function $(z+\varepsilon)^{-\nu/2} K_\nu(\sqrt{z+\varepsilon})$ is CM on $[0, \infty)$ and therefore the function $(\|\mathbf{h}\| + \varepsilon)^{-\nu} K_\nu(\|\mathbf{h}\| + \varepsilon)$ is positive definite on \mathbb{R}^ℓ , for any ℓ . The proof follows by setting $\nu = 1/2$ and $\nu = 1$ corresponding to $d = 3$ and $d = 4$ respectively. \square

We highlight that, even though the functions in Eq. (2.10) have been obtained from (2.9) for specific space dimensions, all covariance functions $C_{d,\varepsilon}(\mathbf{h})$ obtained by application of Proposition 3.1 are valid on \mathbb{R}^ℓ , $\ell \in \mathbb{N}^*$, for any value $\varepsilon > 0$.

3.1.2. Variograms of regularized BG random fields

The functions $C_{d,\varepsilon}(\mathbf{h})/C_d(\varepsilon)$ are valid correlation functions for any value $\varepsilon > 0$. The associated normalized variograms are

$$\tilde{\gamma}_{d,\varepsilon}(\mathbf{h}) = \begin{cases} 1 - \frac{K_0(r/a+\varepsilon_a)}{K_0(\varepsilon_a)}, & d = 2, \\ 1 - \frac{\varepsilon_a}{r/a+\varepsilon_a} \exp(-r/a), & d = 3, \\ 1 - \frac{\varepsilon_a}{r/a+\varepsilon_a} \frac{K_1(r/a+\varepsilon_a)}{K_1(\varepsilon_a)} & d = 4 \end{cases} \quad (3.1)$$

with $\varepsilon_a = \varepsilon/a$. In Figure 1, several 2D normalized variograms $\tilde{\gamma}_{2,\varepsilon}$ are shown for various values of a and ε_a . The parameter ε_a has a significant effect on the shape of the variogram, which we next study analytically for $\varepsilon_a \ll 1$. Note that larger values of ε_a also provide valid variograms, but in this case the function $C_{d,\varepsilon}(r)$ can barely be considered as an approximation of the Boltzmann-Gibbs model explored in this work. At the end of this section, we further show that the $C_{d,\varepsilon}$ tends to the exponential model as $\varepsilon \rightarrow \infty$.

Let us first consider the case $d = 2$, with $\varepsilon_a \ll 1$. For the sake of simplifying notation we drop the index d . The dependence of the variogram at very short distances can be obtained from the *leading-order approximation* $K_0(z) \simeq -\ln(z/2) - \gamma_E$, where γ_E is the Euler-Mascheroni constant. This approximation is very accurate if $z < 0.1$, and it remains practically useful as long as $z < 0.2$. Direct calculation of Eq. (3.1) using the approximation yields the normalized variogram at distance ε :

$$\tilde{\gamma}_\varepsilon(\varepsilon) \simeq -\frac{\ln(2)}{\ln(\varepsilon_a/2) + \gamma_E}. \quad (3.2)$$

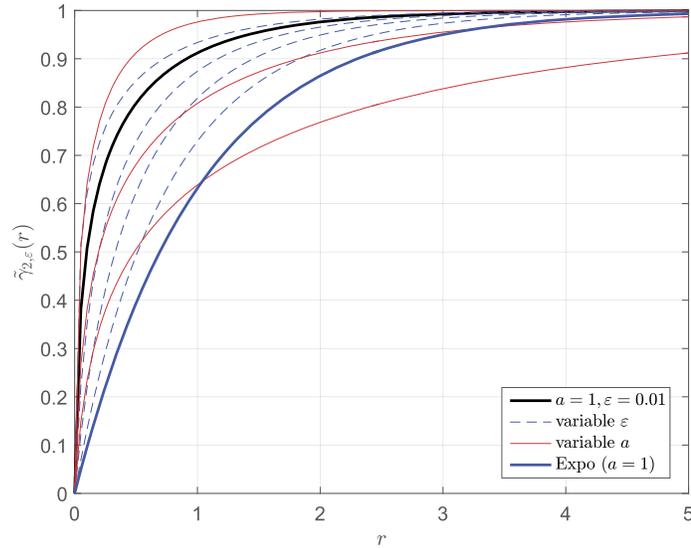


FIG 1. Theoretical normalized variogram $\tilde{\gamma}_{2,\varepsilon}(r)$ for various values of the parameters a and ε . Black line: $a = 1$ and $\varepsilon = 0.01$. Blue dashed lines: $a = 1$ and $\varepsilon = 1/R$ with (from top to bottom) $R = 500, 20, 5, 1$. Blue continuous line: exponential variogram with range $a = 1$. Red continuous lines: $\varepsilon/a = 0.01$ for $a = 0.5, 2, 5$ (a increases from top to bottom).

Furthermore, given a specific level $\tilde{\gamma}_\varepsilon(r_g) = g \ll 1$, based on Eq. (3.2) the distance r_g where this level is achieved is given by

$$r_g/a = 2^{1-1/g} e^{-\gamma_E}.$$

Let us now compute the distance $r_{1/2}$ such that $\tilde{\gamma}_\varepsilon(r_{1/2}) = 1/2$. Using the leading-order approximation of $K_0(\cdot)$, it follows that

$$\tilde{\gamma}_\varepsilon(r_{1/2}) \simeq 1 - \frac{\ln((r_{1/2} + \varepsilon)/2a) + \gamma_E}{\ln(\varepsilon/2a) + \gamma_E},$$

based on which the distance $r_{1/2}$ is given by

$$\varepsilon + r_{1/2} \simeq \sqrt{2ae^{-\gamma_E}} \sqrt{\varepsilon}.$$

If we assume that $\varepsilon \ll r_{1/2}$, one finally gets

$$r_{1/2}/a = \sqrt{2e^{-\gamma_E}} \sqrt{\varepsilon_a}, \tag{3.3}$$

for which the condition $\varepsilon \ll \sqrt{\varepsilon}$ is indeed verified. These computations show that the behavior of the normalized variogram at very short distances is controlled by $\varepsilon_a = \varepsilon/a$, see Eq. (3.2). Furthermore, the distance at which the normalized variogram reaches 1/2 is proportional to the range and proportional to $\sqrt{\varepsilon_a}$. The *practical range* of the variogram can be defined as the distance r_η such that

$\tilde{\gamma}_\varepsilon(r_\eta) = 1 - \eta$, with some $\eta \ll 1$. According to Eq. (3.1), this is given by the solution of the equation $K_0(r_\eta/a + \varepsilon_a) = \eta K_0(\varepsilon_a)$, which involves a nonlinear combination of a and ε_a .

Let us now examine the variogram corresponding to $d = 3$. Here again we drop the subscript $d = 3$ for the sake of lighter notation. Then,

$$\tilde{\gamma}_\varepsilon(\varepsilon) = 1 - \frac{e^{-\varepsilon/a}}{2} \simeq \frac{1}{2} + \frac{\varepsilon_a}{2}.$$

Notice that this value is very close to $1/2$ when $\varepsilon_a \ll 1$, which is usually the case. For $d = 4$, the first order Taylor series expansion $K_1(z) \simeq 1/z$ as $z \rightarrow 0$ leads to

$$\tilde{\gamma}_\varepsilon(\varepsilon) \simeq 1 - \frac{\varepsilon_a}{2\varepsilon_a} \frac{\varepsilon_a}{2\varepsilon_a} = \frac{3}{4},$$

independently on a . These findings imply that as d increases, the (regularized) variograms $\tilde{\gamma}_{d,\varepsilon}$ show an increasingly steep behavior at short distances, and the associated BG random fields have increasingly rough sample paths.

Next, we consider the case $\varepsilon_a \gg 1$, i.e., the asymptotic behavior as $\varepsilon_a \rightarrow \infty$. When $d = 3$, one gets immediately

$$\gamma_{3,\varepsilon}(r) = 1 - e^{-r/a} \frac{\varepsilon_a}{\varepsilon_a + r/a}.$$

For $d = 2$ and $d = 4$, following Abramowitz and Stegun [1, Eq. 9.7.2], as $z \rightarrow \infty$ we obtain the approximation

$$K_\nu(z) = \left(\frac{\pi}{2z}\right)^{1/2} e^{-z} \left(1 + \frac{4\nu^2 - 1}{8z} + \frac{(4\nu^2 - 1)(4\nu^2 - 9)}{128z^2} + \dots\right).$$

Hence, as $\varepsilon \gg a$, and considering $\nu = d/2 - 1$, one gets that

$$\tilde{\gamma}_{d,\varepsilon}(r) = 1 - \left(\frac{\varepsilon_a}{\varepsilon_a + r/a}\right)^\nu \frac{K_\nu(\varepsilon/a + r/a)}{K_\nu(\varepsilon/a)} \simeq 1 - e^{-r/a} \left(\frac{\varepsilon_a}{\varepsilon_a + r/a}\right)^{\nu+1/2}.$$

Therefore, in all cases $\tilde{\gamma}_{d,\varepsilon}$ tends to the exponential variogram as ε_a tends to infinity.

In conclusion, the parameter ε_a plays the role of a regularity parameter that controls the shape of the covariance function, ranging from a generalized covariance function with a slope at the origin proportional to ε_a^{-1} if $\varepsilon_a \ll 1$, to an exponential model if $\varepsilon_a \gg 1$.

3.2. Frequency-space regularization

The divergent covariance functions in Eq. (2.10) (for $d = 2, 3, 4$) can also be regularized by imposing a *frequency cutoff* $\Lambda > 0$, i.e., by assuming that the spectral density is given by

$$\tilde{C}_\Lambda(\mathbf{k}) = \frac{\mathbb{I}(\|\mathbf{k}\| \leq \Lambda)}{\alpha_0 + \alpha_1 \|\mathbf{k}\|^2}. \tag{3.4}$$

Note that the *SLI spectral density* in Eq. (2.6) is given by $\tilde{C}(\mathbf{k}) = \lim_{\Lambda \rightarrow \infty} \tilde{C}_\Lambda(\mathbf{k})$. Using the frequency cutoff $\Lambda < \infty$, the variance is given by the integral

$$\sigma_\Lambda^2 = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{C}_\Lambda(\mathbf{k}) d\mathbf{k}, \tag{3.5}$$

which takes a finite value for $d \geq 2$.

The finite cutoff replaces the upper integration limit in the spectral integral of Eq. (2.8) for the covariance function. To our knowledge, this integral with finite upper limit cannot be explicitly evaluated for every $r \geq 0$. However, for large r the equations (2.10) provide a very accurate approximation of the regularized covariance. A finite value of Λ implies (i) that the regularized variance (3.5) is also finite and (ii) that the random field is mean-square differentiable for every order. Hence, the variogram function is well-defined for $r \rightarrow 0$, and in addition $\gamma(r) = c_0 r^2 + O(r^4)$.

Indeed, using the spectral representation (2.8), we obtain the following representation for finite Λ and for $d \geq 2$:

$$\gamma_\Lambda(r) = \frac{1}{(2\pi)^{d/2} \alpha_0} \left[\lim_{r \rightarrow 0} \int_0^\Lambda \frac{k^{\nu+1} J_\nu(rk)}{r^\nu (1+a^2 k^2)} dk - \int_0^\Lambda \frac{k^{\nu+1} J_\nu(rk)}{r^\nu (1+a^2 k^2)} dk \right]. \tag{3.6}$$

To approximate the above expression, we use the Taylor series expansion of the Bessel function $J_\nu(\cdot)$, which is given by [1, Eq. 9.1.10]

$$J_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^\infty \frac{(-x^2/4)^k}{k! \Gamma(\nu+k+1)} = \left(\frac{x}{2}\right)^\nu \left[1 - \frac{x^2}{4\Gamma(\nu+2)} + O(x^4) \right]. \tag{3.7}$$

In light of the above, the expansion of the variogram function around $r = 0$ is given by

$$\begin{aligned} \gamma_\Lambda(r) &= \frac{1}{(2\pi)^{d/2} \alpha_0} \left[\int_0^\Lambda \frac{k^{2\nu+1} dk}{2^\nu (1+a^2 k^2)} - \int_0^\Lambda \frac{k^{2\nu+1} \left(1 - \frac{k^2 r^2}{4\Gamma(\nu+2)}\right)}{2^\nu (1+a^2 k^2)} dk \right] + O(r^4) \\ &= \frac{r^2}{2^{d+1} \pi^{d/2} \Gamma(\frac{d}{2} + 1) \alpha_0} \int_0^\Lambda \frac{k^{d+1}}{1+a^2 k^2} dk + O(r^4). \end{aligned}$$

This integral is evaluated using the change-of-variable transformation $ak \rightarrow x$ and leads to the general expression

$$\gamma_\Lambda(r) = \frac{B_d(a\Lambda)}{\alpha_0 a^d} \left(\frac{r^2}{a^2}\right) + O(r^4) = \sigma^2 B_d(a\Lambda) \left(\frac{r^2}{a^2}\right) + O(r^4), \tag{3.8a}$$

where $B_d(a\Lambda)$ is a dimensionless constant. More precisely for $d = 2, 3, 4$, we obtain

$$B_d(a\Lambda) = \begin{cases} \frac{1}{16\pi} [(a\Lambda)^2 - \ln(1+a^2\Lambda^2)], & d = 2, \\ \frac{1}{48\pi^{3/2} \Gamma(5/2)} [(a\Lambda)^3 - 3a\Lambda + 3 \tan^{-1}(a\Lambda)], & d = 3, \\ \frac{1}{256\pi^2} [(a\Lambda)^4 - 2(a\Lambda)^2 + 2 \ln(1+a^2\Lambda^2)], & d = 4. \end{cases} \tag{3.8b}$$

The variance of the spectrally regularized BG process with spectral density (2.6) is given by

$$\sigma_\Lambda^2 = \frac{S_d}{(2\pi)^d} \int_0^\Lambda \frac{k^{d-1}}{a_0(1+a^2k^2)} dk.$$

Evaluation of the above integral leads to the following expressions:

$$\sigma_\Lambda^2 = \begin{cases} \frac{S_2}{(2\pi)^2} \frac{\ln(1+a^2\Lambda^2)}{2\alpha_0 a^2} & = \frac{\sigma^2}{4\pi} \ln(1+a^2\Lambda^2), & d = 2, \\ \frac{S_3}{(2\pi)^3} \frac{a\Lambda - \tan^{-1}(a\Lambda)}{\alpha_0 a^3} & = \frac{\sigma^2}{2\pi^2} [a\Lambda - \tan^{-1}(a\Lambda)], & d = 3, \\ \frac{S_4}{(2\pi)^4} \frac{(a\Lambda)^2 - \ln(1+a^2\Lambda^2)}{2\alpha_0 a^4} & = \frac{\sigma^2}{16\pi^2} [(a\Lambda)^2 - \ln(1+a^2\Lambda^2)], & d = 4. \end{cases} \tag{3.9}$$

Thanks to the high frequency cutoff, the variance is finite as can be seen from (3.9), and the variogram is therefore bounded. Thus, a normalized version of the variogram can be defined, which we denote by $\tilde{\gamma}_\Lambda(r) = \gamma_\Lambda(r)/\sigma_\Lambda^2$. Let us consider the case $d = 2$. Based on the spectral representation Eq. (3.6), the series expansion Eq. (3.7) of the Bessel function J_0 , and the regularized variance Eq. (3.9), the series expansion of the *normalized variogram* becomes

$$\tilde{\gamma}_\Lambda(r) = -\frac{2a^2}{\ln(1+a^2\Lambda^2)} \sum_{k=1}^\infty \frac{(-1)^k}{(k!)^2} \left(\frac{r}{2}\right)^{2k} u_k, \tag{3.10a}$$

with the series coefficients u_k being defined by the integrals

$$u_k = \int_0^\Lambda \frac{x^{2k+1}}{1+a^2x^2} dx = \frac{1}{a^2} \int_0^\Lambda x^{2k-1} dx - \frac{1}{a^2} u_{k-1} = \frac{\Lambda^{2k}}{a^2} \left(\frac{1}{2k} - \frac{u_{k-1}}{\Lambda^{2k}}\right). \tag{3.10b}$$

Let us define $\tilde{u}_k = u_k/\Lambda^{2k}$ with $\tilde{u}_0 = \ln(1+a^2\Lambda^2)/2a^2$. Then, the u_k are obtained from the following recursive equation

$$\tilde{u}_k = \frac{1}{2k} - \frac{\tilde{u}_{k-1}}{\Lambda^2},$$

which shows that \tilde{u}_k tends rapidly to $1/(2k)$ for increasing k . Finally, the following power series is obtained for $\tilde{\gamma}_\Lambda(r)$

$$\tilde{\gamma}_\Lambda(r) = -\frac{2a^2}{\ln(1+a^2\Lambda^2)} \sum_{k=1}^\infty \left[-\left(\frac{r\Lambda}{2}\right)^2\right]^k \frac{\tilde{u}_k}{(k!)^2}. \tag{3.11}$$

The power series expansion in Eq. (3.11) converges absolutely for all r . This can be shown by means of D’Alembert’s ratio test [46], thanks to the factorial term $(k!)^2$ in the denominator. However, the expansion is not very useful in practice, since the rapid growth of the term $(r\Lambda/2)^{2k}$ with increasing r makes the series unstable as soon as $r\Lambda/2 > 20$. A more stable approach is the expansion of the Hankel transform (2.8), which makes use of the roots of the zero-order Bessel function of the first kind and leads to semi-analytical expressions for $d = 2, 4$ [4, 22]. For numerical computation of $\gamma_\Lambda(r)$ by means of Eq. (3.6), standard numerical integration routines provide stable results (e.g., using the `integrate` routine in the base library of the R software).

3.3. Matching the regularization schemes

In the preceding two subsections, we presented real-space and frequency-space regularization methods for the generalized covariance functions (2.10) in $d = 2, 3, 4$. Both approaches provide expressions that yield finite variance values, while maintaining the form of the generalized covariances in (2.10) at larger lags. Hence, it is useful to investigate conditions on the regularization parameters for which both approaches lead to the same variance, i.e., $\sigma_\varepsilon^2 = C_\varepsilon(0) = C(\varepsilon) = \sigma_\Lambda^2$. To better understand this constraint, consider the fact that the variance is given by the integral of the spectral density as evidenced in (3.5). Both regularization schemes modify the spectral density (3.4) at high frequencies ($\|\mathbf{k}\| \gg 0$). Hence, the matching of the variance ensures that the integrals of the respective spectral densities, $\tilde{C}_\Lambda(\|\mathbf{k}\|)$ and $\tilde{C}_\varepsilon(\|\mathbf{k}\|)$, are the same over the entire range of $\|\mathbf{k}\|$, and thus over the regime of high $\|\mathbf{k}\|$ as well.

By taking into account (2.10) and (3.9), the *variance matching* condition leads to the following equations:

$$\begin{aligned} \frac{\sigma^2}{2\pi} K_0(\varepsilon/a) &= \frac{\sigma^2}{4\pi} \ln(1 + a^2\Lambda^2), & d = 2 \\ \frac{\sigma^2}{4\pi} \frac{\exp(-\varepsilon/a)}{\varepsilon/a} &= \frac{\sigma^2}{2\pi} [a\Lambda - \tan^{-1}(a\Lambda)], & d = 3, \\ \frac{\sigma^2}{4\pi^2} \frac{K_1(r/a)}{r/a} &= \frac{\sigma^2}{16\pi^2} [(a\Lambda)^2 - \ln(1 + a^2\Lambda^2)], & d = 4. \end{aligned}$$

We further simplify the equations above by considering that in order to only modify the behavior of the covariance near zero, it must hold that $a\Lambda \gg 1$ and $\varepsilon/a \ll 1$. We therefore use the following approximations: $K_0(\varepsilon/a) \simeq -\gamma_E - \ln(\varepsilon/2a)$, $K_1(\varepsilon/a) \simeq a/\varepsilon$, $\ln(1 + a^2\Lambda^2) \simeq 2\ln(a\Lambda)$, $\exp(-\varepsilon/a) \simeq 1$, and $a\Lambda - \tan^{-1}(a\Lambda) \simeq a\Lambda$. Then, the variance matching condition simply becomes $\Lambda\varepsilon = 2e^{-\gamma_E}$ in $d = 2$, $\Lambda\varepsilon = \pi/2$ in $d = 3$, and $\Lambda\varepsilon = 2$ in $d = 4$. As a first order approximation one could thus set $\Lambda\varepsilon \simeq d/2$. In addition to the assumptions $a\Lambda \gg 1$ and $\varepsilon/a \ll 1$, the above analysis demonstrates that the variance matching condition also requires $\Lambda\varepsilon \sim O(1)$. These conditions allow a wide range of combinations of ε and Λ values.

A different approach for matching the regularization schemes involves matching the short distance behavior of the variogram function. As before, we assume $a\Lambda \gg 1$ and $\varepsilon/a \ll 1$. In $d = 2$, we consider the first term of the series expansion (3.8) for the normalized variogram $\tilde{\gamma}_\Lambda(r)$, which is given by

$$\left(\frac{r\Lambda}{2}\right)^2 \left[\frac{1}{\ln(1 + a^2\Lambda^2)} - \frac{1}{a^2\Lambda^2} \right].$$

Evaluating $\tilde{\gamma}_\Lambda(r)$ at $r = \varepsilon$, we obtain

$$\tilde{\gamma}_\Lambda(\varepsilon) \simeq \left(\frac{\varepsilon\Lambda}{2}\right)^2 \left[\frac{1}{\ln(1 + a^2\Lambda^2)} - \frac{1}{a^2\Lambda^2} \right] \simeq \left(\frac{\varepsilon\Lambda}{2}\right)^2 \frac{1}{2\ln(a\Lambda)},$$

where the last approximation is based on $a\Lambda \gg 1$. If we assume that $\Lambda\varepsilon = b$ for some fixed scalar b , then

$$\tilde{\gamma}_\Lambda(\varepsilon) \simeq -\frac{b^2/8}{\ln(\varepsilon/a) - \ln b}.$$

On the other hand, the variogram at distance ε according to real-space regularization is given by (3.2), $\tilde{\gamma}_\varepsilon(\varepsilon) \simeq -\ln(2)/(\ln(\varepsilon/a) - \ln(2) + \gamma_E)$. To match the two regularized variograms at distance ε , we numerically determine the locus $b(\varepsilon)$ of the nonlinear equation $\tilde{\gamma}_\Lambda(\varepsilon) - \tilde{\gamma}_\varepsilon(\varepsilon) = 0$, for $\log(\varepsilon/a) \in [-10, -2]$. Figure 2 plots ε/a against the respective values b . As evidenced in this plot, the two variograms match for $b \in [2.45, 2.75]$ and $\varepsilon/a \leq 0.1$, which then implies that $\Lambda \gtrsim 10$. Hence, we conclude that both matching approaches lead to very similar results, despite their different origins.

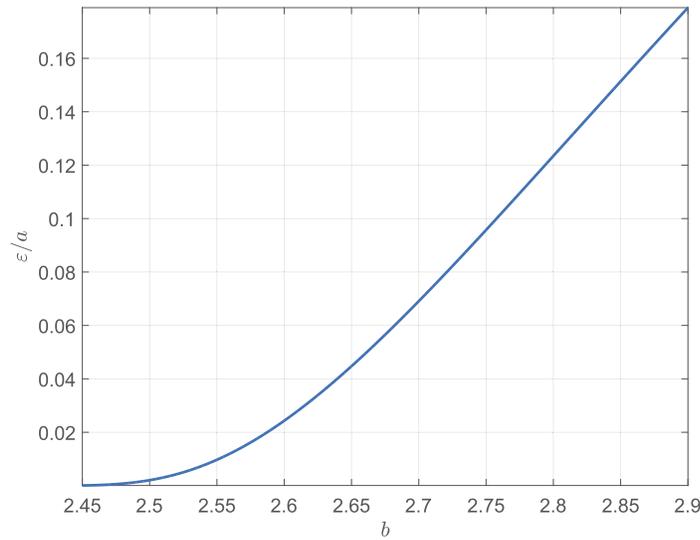


FIG 2. Plot of the curve representing the zero-level locus of the regularized (normalized) variogram difference $\tilde{\gamma}_\Lambda(\varepsilon) - \tilde{\gamma}_\varepsilon(\varepsilon)$ in the parameter space (b, ε_a) .

4. Statistical models for lattice data

In this section, we present a discretization of the continuum BG model over a regular grid in dimension d . We consider a hyper-rectangular grid $\mathcal{G} = \{1, \dots, n_1\} \times \dots \times \{1, \dots, n_d\} \subset \mathbb{Z}^d$ with a hyper-cubic unit cell. The d -dimensional vectors \mathbf{s}_k , $k = 1, \dots, n$ with $n = \prod_{i=1}^d n_i$ are the grid node locations, and $\mathbf{x} = (x_1, \dots, x_n)^\top$ with $x_k = x(\mathbf{s}_k)$ is the vector of the field values on the grid. We further allow the grid step $b > 0$ to be different from 1, such that we use the grid $\mathcal{G}_b = \{b\mathbf{j} \mid \mathbf{j} \in \mathcal{G}\}$, where $\mathbf{j} = \mathbf{j}(\mathbf{s}) = (j_1, \dots, j_d)^\top$ is a \mathcal{G} -valued

index vector with one index for each of the d dimensions. We represent grid points \mathbf{s} as follows: $\mathbf{s} = b \sum_{i=1}^d j_i \mathbf{e}_i$, where $\{\mathbf{e}_i\}_{i=1}^d$ is the set of unit vectors $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ with 1 at the i -th position, and \mathbf{j} is the \mathcal{G} -valued index defined above.

4.1. Energy terms for lattice models

By using a discrete approximation of the gradient evaluated at the grid points in \mathcal{G}_b , the energy term (2.1) can be expressed as a bilinear functional of $\mathbf{x} : \mathcal{G} \rightarrow \mathbb{R}^n$. Using the forward difference approximation for the gradient, i.e.,

$$\nabla_i x(\mathbf{s}) = \frac{x(\mathbf{s} + b\mathbf{e}_i) - x(\mathbf{s})}{b}, \quad i = 1, \dots, d,$$

the discretized contribution of a grid point $\mathbf{s} \in \mathcal{G}_b$ to the integral in the energy functional of the Boltzmann-Gibbs model (2.1) is given by

$$\frac{1}{2b^d} \left[\alpha_0 x(\mathbf{s})^2 + \alpha_1 b^{-2} \left(dx(\mathbf{s})^2 - 2 \sum_{i=1}^d x(\mathbf{s})x(\mathbf{s} + b\mathbf{e}_i) + \sum_{i=1}^d x(\mathbf{s} + b\mathbf{e}_i)^2 \right) \right],$$

where the leading term corresponds to the normalization with respect to the hyper-volume b^d of the grid cell containing \mathbf{s} . Since each point $\mathbf{s} \in \mathcal{G}$ is the forward neighbor of the d points $\mathbf{s} - b\mathbf{e}_i$, $i = 1, \dots, d$, the overall additive contribution (to the approximation of the energy integral) of terms that directly involve $x_k = x(\mathbf{s}_k)$ can be summarized as follows:

$$\mathcal{I}_k = \frac{1}{2b^d} \left[\alpha_0 x_k^2 + \alpha_1 b^{-2} \left(2dx_k^2 - 2 \sum_{i=1}^d x_k x(\mathbf{s}_k + b\mathbf{e}_i) - 2 \sum_{i=1}^d x_k x(\mathbf{s}_k - b\mathbf{e}_i) \right) \right].$$

Therefore, we use the approximation $\mathcal{H}_b(\mathbf{x}; \boldsymbol{\theta}) = \sum_{k=1}^n \mathcal{I}_k$ for the lattice model over \mathcal{G}_b . The points \mathbf{s} at the boundary of \mathcal{G}_b merit special attention. If \mathbf{s} has index vector \mathbf{j} and $j_i = n_i$ for some $i \in \{1, \dots, d\}$, then the forward neighbor $\mathbf{s} + b\mathbf{e}_i$ is part of \mathcal{G}_b . Similarly, if $j_i = 1$ for some $i \in \{1, \dots, d\}$, then \mathbf{s} is forward neighbor of the point $\mathbf{s} - b\mathbf{e}_i$ that is not part of \mathcal{G}_b . In the following equations, we therefore keep the points lying outside the grid \mathcal{G}_b if they are direct neighbors of points within the grid. In statistical practice, even if the field values for such points outside the grid are not observed, it is possible to impose appropriate boundary conditions for such points. A first possibility would be to use Dirichlet boundaries, i.e., to set $x(\mathbf{s}) = 0$ if $\mathbf{s} \notin \mathcal{G}_b$. Another possibility would be to use Neumann boundaries, i.e., to set $x(\mathbf{s} + b\mathbf{e}_i) = x(\mathbf{s})$ if exactly one of the two points \mathbf{s} or $\mathbf{s} + b\mathbf{e}_i$ is not part of \mathcal{G}_b .

The energy term $\mathcal{H}_b(\mathbf{x}; \boldsymbol{\theta})$ discretized over the grid \mathcal{G}_b is bilinear with respect to components of \mathbf{x} , such that it can be written in matrix notation as follows:

$$\mathcal{H}_b(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{2} \mathbf{x}^\top \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x}. \quad (4.1)$$

Since the original energy functional and the discrete approximation in (4.1) are non-negative for any $\alpha_0, \alpha_1 \geq 0$, such that $\max(\alpha_0, \alpha_1) > 0$, the precision matrix $\mathbf{Q}(\boldsymbol{\theta})$ is *non-negative definite*. Moreover, the precision matrix is *positive definite* if at least one of α_0, α_1 is positive, and if in addition $\max_{i=1}^d n_i \geq 2$ in the case of Neumann boundary conditions. The energy functional (4.1) tends to the continuum energy (2.1) at the limit of infill asymptotics $b \rightarrow 0$ by considering the grid where the number of grid points n_i is replaced by $\lceil n_i/b \rceil$ in each dimension $i = 1, \dots, d$.

The precision matrix can be additively decomposed as follows:

$$\mathbf{Q}(\boldsymbol{\theta}) = \alpha_0 b^{-d} \mathbf{I}_n + \alpha_1 b^{-(d+2)} \mathbf{Q}_1, \tag{4.2a}$$

where \mathbf{I}_n is the $n \times n$ identity matrix which results from the energy of the squared fluctuations, and \mathbf{Q}_1 is the *gradient precision sub-matrix* which incorporates the sum of the squared differences and is given by

$$[\mathbf{Q}_1]_{k_1, k_2} = 2d \mathbb{I}(k_1 = k_2) - \mathbb{I}(\mathbf{s}_{k_2} \in \{\mathbf{s}_{k_1} \pm b\mathbf{e}_i \mid i = 1, \dots, d\}), \tag{4.2b}$$

where $k_1, k_2 \in \{1, \dots, n\}$.

4.2. Links to common lattice models

Using the above expressions for the precision matrix elements, we can explicitly provide the conditional distribution of the observation x_k at a point \mathbf{s}_k given the observations at its grid neighbors $\text{NB}_k = \{\mathbf{s}_k \pm b\mathbf{e}_i \mid i = 1, \dots, d\}$ with $|\text{NB}_k| = 2d$. Standard results for (arbitrary) Gaussian vectors [e.g., 43] imply that the conditional variance is given by $1/[\mathbf{Q}]_{k,k}$, and the conditional mean is given by $-[\mathbf{Q}]_{k,k}^{-1} \sum_{\tilde{k} \neq k} [\mathbf{Q}]_{k,\tilde{k}} x_{\tilde{k}}$, such that the conditional distribution can be written as follows:

$$x_k \mid \mathbf{x}_{-k} \sim \mathcal{N} \left(\frac{\alpha_1}{b^2 \alpha_0 + 2d\alpha_1} \sum_{\tilde{k} \in \text{NB}_k} x_{\tilde{k}}, \frac{b^{d+2}}{b^2 \alpha_0 + 2d\alpha_1} \right), \quad k = 1, \dots, n, \tag{4.3}$$

where \mathbf{x}_{-k} refers to the values of the field over all grid points with the exception of x_k . From (4.3), it follows immediately that our grid approximation of the model is closely related to two widely used models for lattice data in spatial statistics. The most popular and fundamental model for areal data observed over known graphs, which are defined by means of the respective adjacency matrix, is due to Besag [6, 7, 8, 43], and it is often referred to as the *conditionally autoregressive (CAR)* model. The general Besag model can be characterized through its conditional distributions

$$x_k \mid \mathbf{x}_{-k} \sim \mathcal{N} \left(\frac{1}{|\text{NB}_k|} \sum_{\tilde{k} \in \text{NB}_k} x_{\tilde{k}}, \frac{1}{|\text{NB}_k| \tau} \right), \quad k = 1, \dots, n,$$

where $\tau > 0$ is a precision parameter. Therefore, the Boltzmann-Gibbs model characterized by the conditional distributions in (4.3) is equivalent to the Besag

model (with points connected to \mathbf{s} given by $\mathbf{s} \pm b\mathbf{e}_i$, $i = 1, \dots, d$) if $\alpha_0 = 0$ and $\tau = \alpha_1/b^{-(d+2)}$.

An extension of the Besag model is the Leroux model [36], which is characterized by the precision matrix

$$\mathbf{Q}_{\text{Leroux}} = \tau_{\text{Leroux}} [(1 - \lambda)\mathbf{I}_n + \lambda\mathbf{Q}_{\text{Besag}}(1)], \quad \lambda \in [0, 1],$$

where $\tau_{\text{Leroux}} > 0$ is a precision parameter and $\mathbf{Q}_{\text{Besag}}(1)$ denotes the precision matrix of the Besag model with $\tau = 1$; the general Besag model arises as a boundary case when $\lambda = 1$ and $\tau = \tau_{\text{Leroux}}$. The Leroux model interpolates between the Besag model ($\lambda = 1$) and a model with spatial independence ($\lambda = 0$). The Leroux model is equivalent to the Boltzmann-Gibbs model in (4.3) if

$$\lambda = \frac{\alpha_1}{\alpha_0 b^2 + \alpha_1}, \quad \tau_{\text{Leroux}} = \frac{\alpha_0 b^2 + \alpha_1}{b^{d+2}}. \quad (4.4)$$

As a result, the grid approximation of the Boltzmann-Gibbs model with energy given in (4.1) can be interpreted as a reparametrization of the Leroux model using parameters α_0 and α_1 instead of τ_{Leroux} and λ . The grid step b is not considered as a parameter since it is usually fixed before using the model. The grid approximation, potentially combined with appropriate boundary conditions, can be viewed as another regularization technique for the continuous-space model, where smaller values of b will lead to a more faithful approximation.

5. Parameter estimation

We now explore how the estimation of the parameters of a regularized covariance function of the form $\sigma^2 C_{d,\varepsilon}(\mathbf{h}; a)$ can be achieved. Several standard estimation methods are available for Gaussian random fields: Weighted Least Squares (WLS) [12], Full Likelihood (FL) and Composite Likelihood (CL) [10, 9]. Composite likelihoods are products of “smaller” likelihoods defined on certain subsets of the data, such as data pairs, for which the likelihood is easy to compute. One can consider the marginal likelihood, the conditional likelihood or the likelihood of differences (also referred to as increments) of such pairs. We follow [10] and choose to consider all possible differences $\{u_{ij} = x_i - x_j\}_{1 \leq i < j \leq N}$, where $U_{ij}(\omega) = X(\mathbf{s}_i; \omega) - X(\mathbf{s}_j; \omega)$ is the two-point random increment. Compared to a full likelihood, this approach can easily handle very large datasets, which resonates with the motivation for exploring Boltzmann-Gibbs models. Following the definition of the variogram function, i.e. $\gamma_{ij}(\boldsymbol{\theta}) = \frac{1}{2} \text{Var}[X(\mathbf{s}_i; \omega) - X(\mathbf{s}_j; \omega)]$, we obtain $\text{Var}[U_{ij}(\omega)] = 2\gamma_{ij}(\boldsymbol{\theta})$; note that $\gamma_{ij}(\boldsymbol{\theta})$ is abbreviated notation for $\gamma_{d,\varepsilon}(\mathbf{s}_j - \mathbf{s}_i; \boldsymbol{\theta})$, which emphasizes the dependence upon the parameters $\boldsymbol{\theta}$ to be estimated. Under the Gaussian assumption for the random field $X(\mathbf{s}; \omega)$, $U_{ij}(\omega)$ is distributed as a Gaussian random variable $\mathcal{N}(0, 2\gamma_{ij}(\boldsymbol{\theta}))$. Ignoring the spatial dependence between different variables $U_{ij}(\omega)$, the negative weighted composite likelihood (after dropping the constant terms that are irrelevant for parameter

estimation) is given by

$$\text{cl}(\boldsymbol{\theta}) = \sum_{i=1}^N \sum_{j>i}^N w_{ij} \text{cl}_{ij}(\boldsymbol{\theta}),$$

where the pairwise likelihood terms are defined by

$$\text{cl}_{ij}(\boldsymbol{\theta}) = \frac{\log \gamma_{ij}(\boldsymbol{\theta})}{2} + \frac{u_{ij}^2}{4\gamma_{ij}(\boldsymbol{\theta})}. \quad (5.1)$$

The computational cost of calculating $\text{cl}(\boldsymbol{\theta})$ is $\mathcal{O}(N^2)$. It can be significantly reduced by only considering pairs of sites that are closer than a specified distance r_0 , i.e., by setting $w_{ij} = \mathbb{I}(r_{ij} \leq r_0)$. As shown in [10], choosing an appropriate r_0 not only provides a significant gain in computational time but also increases the statistical efficiency, since pairs of observations separated by distances exceeding the correlation range are uninformative for the spatial scale parameter. The associated CL score statistic,

$$\mathbf{CL}^{(1)}(\boldsymbol{\theta}; \omega) = \sum_{i=1}^N \sum_{j>i}^N w_{ij} \frac{\gamma_{ij}^{(1)}(\boldsymbol{\theta})}{2\gamma_{ij}(\boldsymbol{\theta})} \left[1 - \frac{U_{ij}(\omega)^2}{2\gamma_{ij}(\boldsymbol{\theta})} \right], \quad (5.2)$$

is unbiased, i.e., $\mathbb{E}[\mathbf{CL}^{(1)}(\boldsymbol{\theta}; \omega)] = \mathbf{0}$, where $\mathbf{0}$ is the zero vector in a space with dimension equal to the number of components of $\boldsymbol{\theta}$. In Eq. (5.2) $f^{(1)}$ denotes the gradient (i.e., the vector of partial derivatives) of a function f with respect to $\boldsymbol{\theta}$. This result does not depend on distributional assumptions about $X(\mathbf{s}; \omega)$, since for stationary random fields it always holds that $\mathbb{E}[U_{ij}(\omega)^2] = 2\gamma_{ij}(\boldsymbol{\theta})$. CL is therefore expected to provide better estimates than FL when the Gaussian assumption is violated.

6. Case study: Walker Lake data

6.1. Description of the data set

We use the well-known Walker Lake data set [31] to illustrate statistical aspects of the Boltzmann-Gibbs covariance model. It is available from the `gstat` package in R [40]. The Walker Lake data set was derived from digital terrain elevations of the Walker Lake area in western Nevada, US. A rectangular region containing elevation values at 1.95 million regularly spaced points was divided into a coarse-grained 300×260 grid in which each cell contained a 5×5 array of points. The mean and variance of the elevation within each cell were determined. Through a deterministic function of these two values, a synthetic metal concentration V (herein denoted by X for consistency with our random field notation) was derived for each cell of the grid. A complete description of the construction of the Walker Lake data set is given in [31, Appendix A]. Because the complete data set is highly non-stationary, we selected the $[10, 200] \times [10, 100]$ subdomain containing $N = 17,381$ data points, which we deem compatible with a stationarity assumption (see the left-hand side plot of Figure 3).

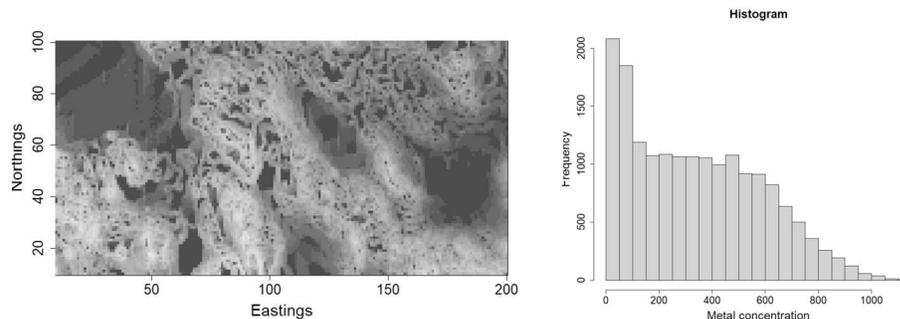


FIG 3. Walker Lake data. Left: Map of synthetic metal concentration in the selected subdomain. Right: Histogram of the concentration values in the subdomain.

6.2. Estimation of the variogram

6.2.1. Estimation with full dataset

We first estimate the parameters of the BG variogram using as training set all 17,381 data in the entire subdomain, based on WLS and CL. Full likelihood is not feasible for numerical reasons due to the large size of the training dataset. As can be seen from the histogram plot on the right-hand side of Figure 3, the concentration follows a non-Gaussian distribution. For numerical optimization we used the R function `nlminb`, which was successfully tested on simulated data in unreported preliminary analyses.

Estimated values of the BG parameter vector $(\sigma^2, a, \varepsilon_a)^\top$ are listed in Table 1, and the estimated variograms are shown on Figure 4. The experimental variogram (black circles) is computed for distances ranging from 0 to 70, using $K = 80$ bins of equal width. When the fitting is performed using WLS, the experimental variogram is closely fitted by the Boltzmann-Gibbs variogram $\sigma^2 \tilde{\gamma}_{2,\varepsilon}(\mathbf{h})$ where $\tilde{\gamma}_{2,\varepsilon}(\mathbf{h})$ is given by Eq. (3.1) (blue continuous line). WLS is used with optimal weights proportional to $n_k / \tilde{\gamma}^2(h_k; \boldsymbol{\theta})$, where h_k is the distance corresponding to bin $k \in \{1, \dots, K\}$ and n_k is the number of pairs in that bin (cf. [12] for a justification of this choice). CL-based variograms estimated with cutoff distances $r_0 = 10$ and $r_0 = 20$, underestimate the variability at distances larger than the respective cutoff distance. For $r_0 = 30$ or larger, the estimated variogram is almost identical to the one estimated with WLS. In all cases, the estimated spatial cutoff parameter verifies $\varepsilon_a \ll 1$ (see Table 1), as discussed in Section 3.

For comparison, an exponential variogram $\sigma^2(1 - \exp(-\|\mathbf{h}\|/a))$ was also fitted with WLS and CL under the same conditions. Its fit to the experimental variogram is evidently inferior to that of the BG model, in particular for intermediate (with respect to the range) distances, approximately between 15 and 35. The additional regularization parameter in the BG variogram allows fitting very accurately the experimental variogram at all distances (short, intermediate

and large). For the exponential variogram, CL systematically underestimates the variability at distances larger than r_0 . The estimates are equal to WLS estimates only if the cutoff distance is equal to 50, which equals half of the North-South diagonal distance in the study domain. From now on, the set of parameters $(\sigma^2, a, \varepsilon_a) = (67, 450, 40.7, 0.07)$ will be considered as the “ground truth”. These parameters correspond to the WLS and CL estimates with $r_0 \geq 30$ (blue curve in Figure 4).

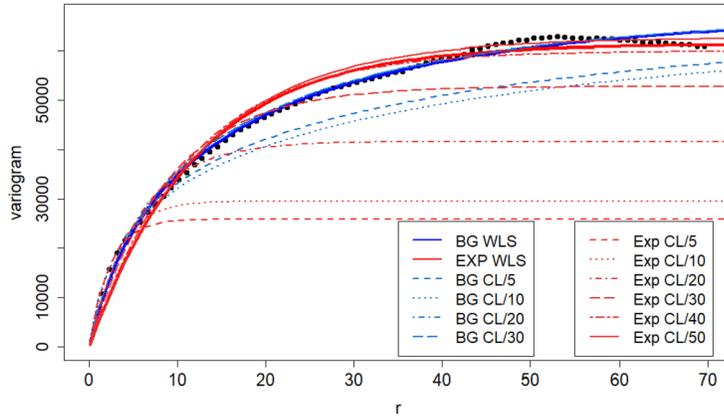


FIG 4. Empirical and estimated theoretical variograms using WLS (continuous lines) and CL (broken lines). Blue: Boltzmann-Gibbs (BG) variograms. Red: Exponential (Exp) variograms.

TABLE 1

Estimates of the Boltzmann-Gibbs and exponential variogram model parameters obtained by means of Composite Likelihood (CL) with several cutoff distances and Weighted Least Squares (last row).

r_0	Boltzmann-Gibbs			Exponential	
	$\hat{\sigma}^2$	\hat{a}	$\hat{\varepsilon}_a$	$\hat{\sigma}^2$	\hat{a}
5	67,450	107.5	0.001	25,875	2.3
10	67,440	133.3	0.007	29,560	2.9
20	67,450	40.7	0.067	41,620	5.6
CL 30	67,450	40.7	0.067	52,780	8.7
40	67,450	40.7	0.067	59,955	11.2
50	67,450	40.7	0.067	62,503	12.5
WLS	67,450	40.7	0.075	61257	12.2

6.2.2. Estimation using smaller training sets

We now assess the ability to estimate the “true” variogram based on training sets of small to moderate size. We will uniformly select 200, 400 and 800 samples from the full subdomain dataset and estimate the parameters using WLS, Full Likelihood (FL) and CL with $r_0 \in \{5, 10, 20, 30, 40, 50\}$. The whole experiment of random subsampling followed by estimation is repeated 100 times.

These estimates are then compared to the “ground truth” established in Section 6.2.1. Since the synthetic metal concentration does not follow the Gaussian distribution, likelihood methods might not perform very well. FL is expected to perform poorly (even if computationally feasible), because it relies heavily on the distributional properties of the multivariate Gaussian distribution.

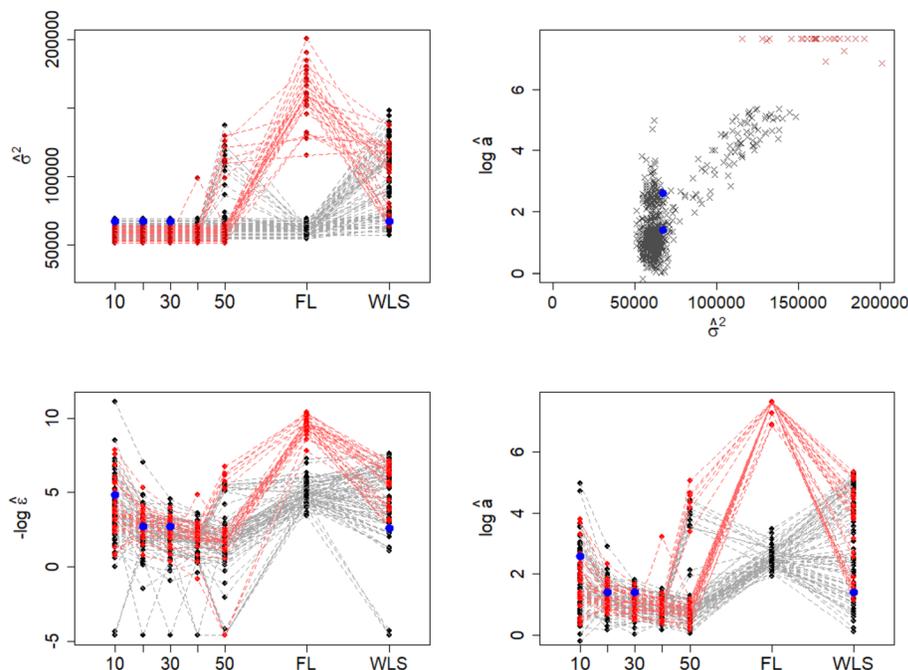


FIG 5. Estimates of BG parameters by means of different estimation methods based on training sets with 400 randomly selected observations (100 repetitions). Estimation methods involve CL with $r_0 = 10, 20, 30, 40, 50$; FL; WLS. Top left and Bottom Row: estimates of σ^2 , $-\log(\varepsilon_a)$ and $\log(\hat{a})$; red points and crosses denote training sets with FL estimates of $\hat{\sigma}^2 \geq 100,000$. Top right: $\log(\hat{a})$ vs. $\hat{\sigma}^2$ for all spatial models. In all plots: blue dots are CL and WLS estimates obtained with full dataset.

Figure 5 shows the estimates of the three BG parameters for different estimation methods as well as the scatter plot of $\log(\hat{a})$ vs. $\hat{\sigma}^2$, for all methods. The estimates are obtained for 100 configurations of $n = 400$ randomly sampled observations. FL leads to a bimodal set of estimates. In the first group, the estimates are close to those obtained with the complete dataset (black markers in Figure 5). However, there is also a second group of estimated values (red markers in Figure 5) with much larger values of $\hat{\sigma}^2$ (from 115,687 to 200,643) and ranges (from 949 to 2,102, with 2,102 representing the upper bound for the range used in the optimization procedure) and smaller ε_a (from 0.0007 to 0.0032). This bimodality is specific to FL in this experiment, even though WLS and CL with large r_0 also lead to estimates that significantly deviate from the

“true” values. Closer inspection reveals that $r_0 \in \{20, 30\}$ consistently leads to the least biased estimates. However, the variance is underestimated due to the small size of the training set. The upper-right window of Figure 5 is a scatter plot of $\hat{\sigma}^2$ and \hat{a} obtained from all the estimation methods (WLS, CL, FL). As evidenced in this plot, $\hat{\sigma}^2$ and \hat{a} are not correlated if $\hat{\sigma}^2$ is approximately around the “true” value. However, for training configurations that lead to larger estimates of $\hat{\sigma}^2$, there is a positive linear correlation between $\log(\hat{a})$ and $\hat{\sigma}^2$.

Very similar patterns were obtained with training sets containing $n = 200$ and $n = 800$ observations. Estimates are more or less dispersed around the “true” values, depending on n . In particular, the FL group of large variance and large range estimates becomes smaller as n increases (and vice-versa, it becomes larger as n decreases). The score statistics associated with CL, given in Eq. (5.2), are very close to zero for all cutoff distances r_0 .

Figure 6 represents the 100 BG model variograms (based on CL parameter estimates with $r_0 = 30$), along with the variograms (experimental and based on CL estimates) obtained from the complete dataset of the Walker Lake subdomain. For the sake of comparison, both Boltzmann-Gibbs and exponential variograms were estimated. The comparison clearly shows that BG estimates are less dispersed and more centered around the “true” (experimental) variogram. This figure provides further evidence that the regularized BG variogram model better fits the spatial dependence of the data than the exponential model, even for moderately sized training sets. Similar plots were obtained for $n = 200$ and $n = 800$ but are not reported herein for the sake of conciseness.

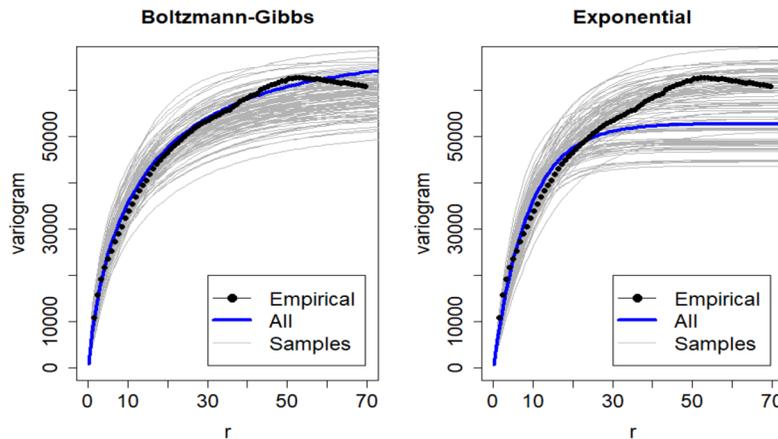


FIG 6. Estimated variograms based on training sets with 400 randomly selected observations (100 repetitions). On both plots, black dots represent the experimental variogram computed with the complete dataset. Blue lines show variograms estimated on the complete dataset, and grey lines correspond to variograms estimated on the 100 training sets, using CL with $r_0 = 30$ in both cases. Left: Boltzmann-Gibbs variograms. Right: Exponential variograms.

6.3. Prediction and cross-validation

To assess the predictive performance of the BG model, the random training sets and their associated parameter estimates obtained in Section 6.2 are now used for prediction (kriging) and cross-validation. For a training set of size n , where $n \in \{200, 400, 800\}$, we use ordinary kriging to predict the synthetic metal concentration X at each of the $N - n$ remaining locations of the Walker Lake subdomain ($N = 17,381$). The partitioning of the dataset into training and validation sets is repeated 100 times. Four different spatial models are used for the comparison. First, ordinary kriging is performed for each randomly selected training set with the BG “ground truth” variogram and the exponential (Exp) variogram (both shown as blue lines in Figure 4). Recall that the “ground truth” parameters were estimated with WLS, and they are listed in Table 1. Then, for each of the above random training sets, the variograms (BG and Exp) with parameters estimated using maximum CL with the optimal cutoff distance (i.e., $r_0 = 30$), are also used for ordinary kriging. The resulting four spatial models are denoted F-BG, F-Exp, S-BG and S-Exp, respectively, where F stands for “Full dataset” (on the selected subdomain) and S for “Subdomain samples”. For each random training set and each spatial model, four different scores are computed: mean error (ME), mean squared error (MSE), mean normalized squared error (MNSE), and linear correlation between the prediction error and the true value (COR). Good predictive performance requires ME close to 0 and MNSE close to one. The latter requirement is due to the fact that (assuming zero bias of the kriging estimates) $\mathbb{E}[(\hat{X}_i - X_i)^2] = \sigma_{K,i}^2$, where X_i and \hat{X}_i are the true value and the kriging estimate at point \mathbf{s}_i , respectively, and $\sigma_{K,i}^2$ is the associated kriging variance. Recall that kriging is the orthogonal projection on the linear subspace spanned by the data, e.g. [12]. Therefore, $\mathbb{E}[(\hat{X}_i - X_i) \hat{X}_i] = 0$, and COR should be as close to 0 as possible. For each of the above four scores and each of the four spatial models, we obtain 100 values associated with the 100 training sets.

The results of the cross-validation analysis are summarized in Table 2. As expected, for all values n and for all the spatial models, kriging is essentially unbiased (with ME being close to zero). Boxplots of the average MSEs are shown in Figure 7. For all n , the MSE is smaller for the BG model than for the Exp model, whether based on the full dataset (F-BG vs. F-Exp) or the subsamples (S-BG vs. S-Exp). The relative MSE improvement of the BG model versus the Exp ranges from 1.5% when $n = 200$ to 2.4% when $n = 800$. When using the Exp model, predictions are slightly more accurate with the “full” model than with the specific model estimated with the sample at hand. The opposite holds when using the BG model, but the difference is very small. A striking result is that NMSE is significantly closer to 1 for the BG than for the Exp model. In other words, the exponential model largely underestimates the magnitude of the kriging error, whereas the BG model has an MNSE score quite close to 1. In addition, COR is always smaller with the BG than for the Exp model. To summarize, the (regularized) BG model provides better scores than the exponential model for all validation metrics considered.

TABLE 2

Assessment of prediction performance of BG and exponential models on the Walker Lake dataset. Sparsity index (p_s), Mean Error (\overline{ME}), Mean Squared Error (\overline{MSE}), Mean Normalized Squared Error (\overline{MNSE}), and mean Correlation between the prediction error and the true value (\overline{COR}), for different training set sizes ($n = 200, 400, 800$) and four different spatial models. Boldfacing is used to denote the optimal values for each score.

n		p_s	\overline{ME}	\overline{MSE}	\overline{MNSE}	\overline{COR}
200	F-BG	0.230	0.94	29,449	1.05	0.016
	F-Exp	0.418	0.91	29,882	1.27	0.068
	S-BG	0.338	0.95	29,633	1.08	0.003
	S-Exp	0.484	0.91	29,974	1.29	0.052
400	F-BG	0.115	-0.31	24,381	1.09	0.033
	F-Exp	0.157	-0.48	24,923	1.43	0.091
	S-BG	0.128	-0.30	24,427	1.06	0.018
	S-Exp	0.180	-0.38	24,870	1.33	0.071
800	F-BG	0.063	0.55	20,486	1.17	0.039
	F-Exp	0.072	0.51	20,985	1.63	0.089
	S-BG	0.063	0.54	20,526	1.19	0.039
	S-Exp	0.073	0.50	20,940	1.54	0.080

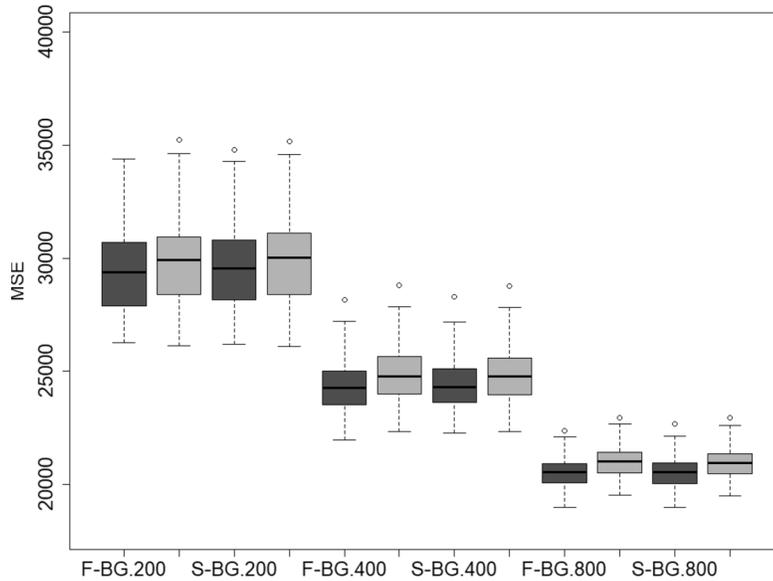


FIG 7. MSE of kriging predictions based on the BG and Exp variogram models for the Walker Lake dataset. Boxplots are shown for all combinations of spatial models and training set sizes based on 100 different partitions of the dataset). First four boxplots (from left to right) correspond to $n = 200$: Full BG (F-BG), Full Exp, Sample BG (S-BG), and Sample Exp. The same order is repeated for $n = 400$ and $n = 800$. See text for details.

A particularly interesting feature of the BG model that enables efficient computation of predictions is the degree of sparsity of the precision matrix. An indicator of sparsity is the proportion of precision matrix entries with an absolute value larger than 10^{-8} , denoted by p_s . The index p_s is listed in the third column of Table 2. Its values show that the precision matrix is always sparser for the BG than for the Exp model. This is a noteworthy finding, since the exponential model possesses a Markov property on the continuous plane [37] and is known to show very sparse precision matrices in practice.

All of the above results of cross-validation analysis provide strong evidence in favor of the regularized Boltzmann-Gibbs model, at least for this dataset.

7. Conclusions

The work presented in this paper was motivated by the strong interest in statistical methodology geared towards spatial models based on physically inspired stochastic differential equations, e.g., in the developments and applications originating in the seminal work of [37], and going further in [11], as well as connections between statistical field theory and spatial models [24, 28, 25]. Starting from the specification of a Boltzmann-Gibbs model that directly and intuitively controls spatial dependence in terms of an energy functional that contains contributions from squares of the fluctuations and their gradients, we have characterized the properties of the resulting random fields in terms of the covariance and the spectrum. In cases where the variance is infinite, we achieve finite covariance structures, useful for simulation and statistical modeling and prediction, by means of regularization schemes in the direct and spectral domains. Specifically, the direct-space covariance regularization introduces a new approximation parameter, $\varepsilon > 0$, which can be estimated jointly with the other model parameters using techniques such as the pairwise difference likelihood. When ε moves from 0 to infinity, the corresponding model smoothly interpolates between the Boltzmann-Gibbs generalized covariance and the exponential covariance. In two dimensions both of the above can be considered as special cases of the widely used Matérn covariance with regularity parameters 0 and 0.5, respectively.

With respect to spectral regularization, instead of using a “hard” (abrupt) frequency cutoff in Eq. (3.4), it is possible to use a tapering function that provides a smooth decay of the spectral density, as proposed in [24]. This is equivalent to a convolution of the random field with a suitable kernel function in direct space; the latter would then act as a filter for the high-frequency fluctuations of the field. Such a choice is certainly appealing due to the smooth decay of the spectral density that it produces. On the other hand, the integrals for the regularized variance (3.5) and variogram function (3.6), will become more complicated.

A grid approximation of the Boltzmann-Gibbs field with appropriate boundary conditions can be viewed as another regularization technique and establishes a direct link to the popular Besag and Leroux lattice models used in spatial statistics. The grid approximation is characterized by sparse precision matrices, which are useful for gridded high-dimensional data thanks to their numerical

benefits. In future work, we plan to develop similar sparse matrix approximations of the Boltzmann-Gibbs models for irregularly spaced locations.

Another promising research direction is the extension of the Boltzmann-Gibbs model to vector random fields, which would be applicable to multivariate continuum data as well as multi-spectral images. In addition, given the simple rational function expression of the Boltzmann-Gibbs spectral density, it is worthwhile to investigate spectral estimation methods for Boltzmann-Gibbs random fields, both for gridded but incomplete and irregularly spaced data. Such approaches based on Whittle's approximation [48] of the Gaussian log-likelihood have been explored in [16, 21].

The Gibbs energy term (2.1) studied herein provides control over local gradients but not directly over the local curvature of the random field. Boltzmann-Gibbs models which include curvature have been developed in [24, 28, 25]. Extensions towards the inclusion of higher-order (than the gradient and curvature) terms allowing for smoother realizations of the Gaussian field, or towards terms defining non-Gaussian Boltzmann-Gibbs fields, are other promising areas of research.

References

- [1] ABRAMOWITZ, M. AND STEGUN, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, Washington, DC, USA. [MR0415962](#)
- [2] ADLER, R. J. AND TAYLOR, J. E. (2009). *Random Fields and Geometry*. Springer Science & Business Media, New York. [MR2319516](#)
- [3] ALBEVERIO, S., MAZZUCCHI, S., AND BRZEZNIAK, Z. (2017). Probabilistic integrals: mathematical aspects. *Scholarpedia* **12**, 5, 10429. Revision #182435.
- [4] ARFKEN, G. B., WEBER, H. J., AND HARRIS, F. E. (2013). *Mathematical Methods for Physicists*, 7th ed. Elsevier, Amsterdam, the Netherlands. [MR0205512](#)
- [5] BERNSTEIN, S. (1929). Sur les fonctions absolument monotones. *Acta Mathematica* **52**, 1–66. [MR1555269](#)
- [6] BESAG, J. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society. Series B (Methodological)* **34**, 1, 75–83. [MR0323276](#)
- [7] BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**, 2, 192–236. [MR0373208](#)
- [8] BESAG, J. AND KOOPERBERG, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82**, 4, 733–746. [MR1380811](#)
- [9] BEVILACQUA, M. AND GAETAN, C. (2015). Comparing composite likelihood methods based on pairs for spatial gaussian random fields. *Statistics and Computing* **25**, 5, 877–892. [MR3375623](#)
- [10] BEVILACQUA, M., GAETAN, C., MATEU, J., AND PORCU, E. (2012). Estimating space and space-time covariance functions for large data sets:

- a weighted composite likelihood approach. *Journal of the American Statistical Association* **107**, 497, 268–280. [MR2949358](#)
- [11] CARRIZO VERGARA, R., ALLARD, D., AND DESASSIS, N. (forthcoming). A general framework for SPDE-based stationary random fields. *Bernoulli*. <http://www.bernoulli-society.org/publications/bernoulli-journal/bernoulli-journal-papers>.
- [12] CHILÈS, J.-P. AND DELFINER, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. Vol. **497**. John Wiley & Sons, Hoboken, NJ, USA. [MR2850475](#)
- [13] DA PRATO, G. AND ZABCZYK, J. (2002). *Second Order Partial Differential Equations in Hilbert Spaces*. Vol. **293**. Cambridge University Press, Cambridge, UK. [MR1985790](#)
- [14] EL KAMEL, J. AND MEHREZ, K. (2018). A function class of strictly positive definite and logarithmically completely monotonic functions related to the modified Bessel functions. *Positivity* **22**, 5, 1403–1417. [MR3863624](#)
- [15] FASSHAUER, G. E. (2007). *Meshfree Approximation Methods with MATLAB*. Interdisciplinary Mathematical Sciences, Vol. **6**. World Scientific, Hackensack, NJ, USA. [MR2357267](#)
- [16] FUENTES, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association* **102**, 477, 321–331. [MR2345545](#)
- [17] GEL'FAND, I. M. AND SHILOV, G. (1968). *Generalized Functions, Volume 2: Spaces of Fundamental and Generalized Functions*. Vol. **261**. Academic Press, New York, NY. [MR0230128](#)
- [18] GLIMM, J. AND JAFFE, A. (1987). *Quantum Physics: A Functional Integral Point of View*, 2nd ed. Springer Science & Business Media, New York, NY. [MR0887102](#)
- [19] GOLDENFELD, N. (1993). *Lectures on Phase Transitions and the Renormalization Group*. Frontiers in Physics. Addison-Wesley, Reading, MA.
- [20] GRADSHTEYN, I. S. AND RYZHIK, I. M. (2007). *Table of Integrals, Series, and Products*, 7th ed. Academic Press, Boston. [MR0669666](#)
- [21] GUINNESS, J. (2019). Spectral density estimation for random fields via periodic embeddings. *Biometrika* **106**, 2 (04), 267–286. [MR3949303](#)
- [22] GUIZAR-SICAÍROS, M. AND GUTIÉRREZ-VEGA, J. C. (2004). Computation of quasi-discrete Hankel transforms of integer order for propagating optical wave fields. *Journal of the Optical Society of America A* **21**, 1, 53–58. [MR2026147](#)
- [23] HIGHAM, N. J. (2015). Regularization. In *The Princeton Companion to Applied Mathematics*, N. J. Higham, M. R. Dennis, P. Glendinning, P. A. Martin, F. Santosa, and J. Tanner, Eds. Princeton University Press, Princeton, NJ, USA, 205–206. [MR3380576](#)
- [24] HRISTOPULOS, D. T. (2003). Spartan Gibbs random field models for geostatistical applications. *SIAM Journal on Scientific Computing* **24**, 6, 2125–2162. [MR2005624](#)
- [25] HRISTOPULOS, D. T. (2015). Covariance functions motivated by spatial random field models with local interactions. *Stochastic Environmental Research and Risk Assessment* **29**, 3, 739–754.

- [26] HRISTOPULOS, D. T. (2020). *Random Fields for Spatial Data Modeling: A Primer for Scientists and Engineers*. Springer Nature, Dordrecht, Netherlands.
- [27] HRISTOPULOS, D. T. AND AGOU, V. D. (2020). Stochastic local interaction model with sparse precision matrix for space-time interpolation. *Spatial Statistics* **40**, 100403. [MR4181140](#)
- [28] HRISTOPULOS, D. T. AND ELOGNE, S. N. (2007). Analytic properties and covariance functions for a new class of generalized Gibbs random fields. *IEEE Transactions on Information Theory* **53**, 12, 4667–4679. [MR2446930](#)
- [29] HRISTOPULOS, D. T. AND PORCU, E. (2014). Multivariate Spartan spatial random field models. *Probabilistic Engineering Mechanics* **37**, 84–92.
- [30] HRISTOPULOS, D. T. AND TSANTILI, I. (2017). Space-time covariance functions based on linear response theory and the turning bands method. *Spatial Statistics* **22**, 2, 321–337. [MR3732853](#)
- [31] ISAAKS, E. H. AND SRIVASTAVA, M. R. (1989). *Applied Geostatistics*. Oxford University Press, New York.
- [32] ISING, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik* **31**, 1, 253–258.
- [33] JACKSON, J. D. (1999). *Classical Electrodynamics*. John Wiley & Sons, Inc., Hoboken, NJ, USA. [MR0436782](#)
- [34] KARDAR, M. (2007). *Statistical Physics of Fields*. Cambridge University Press, Cambridge, UK. [MR2374147](#)
- [35] KLEINERT, H. AND SCHULTE-FROHLINDE, V. (2001). *Critical Properties of Phi-4-theories*. World Scientific, River Edge, NJ. [MR1868655](#)
- [36] LEROUX, B. G., LEI, X., AND BRESLOW, N. (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, M. E. Halloran and D. Barry, Eds. IMA Volumes in Mathematics and its Applications, Vol. **116**. Springer, New York, NY, 179–191. [MR1731684](#)
- [37] LINDGREN, F., RUE, H., AND LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach. *Journal of the Royal Statistical Society, Series B* **73**, 4, 423–498. [MR2853727](#)
- [38] METROPOLIS, N. AND ULAM, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association* **44**, 247, 335–341. [MR0031341](#)
- [39] MONDAL, D. (2015). Applying Dynkin’s isomorphism: an alternative approach to understand the Markov property of the de Wijs process. *Bernoulli* **21**, 3, 1289–1303. [MR3352044](#)
- [40] PEBESMA, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**, 7, 683–691.
- [41] RICHARDS, J. I. AND YOUN, H. K. (1990). *The Theory of Distributions: A Nontechnical Introduction*. Cambridge University Press, New York, NY. [MR1058014](#)
- [42] ROBERT, C. P. AND CASELLA, G. (1999). *Monte Carlo Statistical Methods*. Springer, New York, NY, USA. [MR1707311](#)
- [43] RUE, H. AND HELD, L. (2005). *Gaussian Markov Random Fields: Theory*

- and Applications*. Chapman and Hall/CRC, Boca Raton, FL. [MR2130347](#)
- [44] SCHOENBERG, I. J. (1938). Metric spaces and completely monotone functions. *Annals of Mathematics* **39**, 4, 811–841. [MR1503439](#)
 - [45] SCHWARTZ, L. (1966). *Théorie des Distributions*. Vol. **2**. Hermann, Paris, France. [MR0209834](#)
 - [46] WATSON, G. S. (1964). Smooth regression analysis. *Sankhya Ser. A* **26**, 1, 359–372. [MR0185765](#)
 - [47] WENDLAND, H. (2004). *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics, Vol. **17**. Cambridge University Press, Cambridge, UK. [MR2131724](#)
 - [48] WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 3/4, 434–449. [MR0067450](#)
 - [49] YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions I*. Springer Verlag, New York. [MR0915557](#)